CUk-H02378.63-P011710

AMERICAN JOURNAL OF MATHEMATICS

63

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

R. BRAUER
UNIVERSITY OF MICHIGAN

L. M. GRAVES
UNIVERSITY OF CHICAGO

D. C. LEWIS, JR.
THE JOHNS HOPKINS UNIVERSITY

H. WHITNEY
HARVARD UNIVERSITY

A. WINTNER
THE JOHNS HOPKINS UNIVERSITY

WITH THE COÖPERATION OF

L. BERS
J. L. DOOB
E. K. HAVILAND
N. JACOBSON
A. D. WALLACE

L. V. AHLFORS
R. ARENS
J. D. HILL
N. H. McCOY
J. L. SYNGE

C. B. ALLENDOERFER
R. BAER
P. HARTMAN
J. J. STOKER
O. SZÁSZ

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

510.5
63

VOLUME LXXI

1949

CALCUTTA UNIVERSITY
CENTRAL LIBRARY

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

916710

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

INDEX

PAGE

Arens, Richard. Approximation in, and representation of, certain Banach algebras, ... 763

Baer, Reinhold. Extension types of Abelian groups, ... 461

Baer, Reinhold. Free sums of groups and their generalizations. An analysis of the associative law, ... 706

Bell, James H. Left associates of monic matrices, with an application to unilateral matrix equations, ... 249

Bell, P. O. Differential geometry of a general motion in the complex projective line, ... 427

Bochner, S. and K. Chandrasekharan. Gauss summability of trigonometric integrals, ... 50

Borg, Göran. On a Liapounoff criterion of stability, . . . 67

Borofsky, S. Characterization of a field by a single operation, . 92

Brauer, Alfred and Nathaniel Macon. On the approximation of irrational numbers by the convergents of their continued fractions, ... 349

Busemann, Herbert. The isoperimetric problem for Minkowski area, 743

Chandrasekharan, K. See Bochner and Chandrasekharan, page 50.

Chang, Shih-Hsun. A class of integro-differential equations, . . 563

Chow, Wei-Liang. On compact complex analytic varieties, . . 893

Clifford, A. H. Semigroups without nilpotent ideals, . . . 834

Coddington, Earl A. and Aurel Wintner. On the classical existence theorem of analytic differential equations, . . . 886

Cohn, Harvey. Some diophantine aspects of modular functions. I. Essential singularities, ... 403

Cowling, V. F. Analytic continuation of factorial series, . . . 283

De Bruijn, N. G. The asymptotically periodic behavior of the solutions of some linear functional equations, ... 313

Doss, Raouf. On uniform spaces with a unique structure, . . 19

Everett, C. J. and G. Whaples. Representations of sequences of sets, 287

Fort, Tomlinson. Quasi-monotone series, ... 227

Gilbarg, David. A characterization of non-isentropic irrotational flows, 687

Haimo, Franklin. A class of inverse limit groups, ... 171

Hartman, Philip. On the spectra of slightly disturbed linear oscillators, ... 71

Hartman, Philip and Aurel Wintner. A criterion for the non-degeneracy of the wave equation, ... 206

Hartman, Philip and Aurel Wintner. On the location of spectra of wave equations, ... 214

PAGE

Hartman, Philip and Aurel Wintner. On the Laplace-Fourier transcendents, ... 367

Hartman, Philip and Aurel Wintner. Oscillatory and non-oscillatory linear differential equations, ... 627

Hartman, Philip and Aurel Wintner. A separation theorem for continuous spectra, ... 650

Hartman, Philip and Aurel Wintner. On the classical existence theorem of linear differential equations, ... 859

Hartman, Philip and Aurel Wintner. Separation theorems for bounded Hermitian forms, ... 865

Hartman, Philip. A characterization of the spectra of one-dimensional wave equations, ... 915

Herstein, Israel Nathan. Divisor algebras, ... 800

Hochschild, G. Double vector spaces over division rings, . . 443

Hsiung, Chuan-Chih. Invariants of intersection of certain pairs of curves in N-dimensional space, ... 678

Hsiung, Chuan-Chih. Affine invariants of a pair of hypersurfaces, . 879

Hu, Sze-tsen. Structure of the homotopy groups of mapping spaces, 574

Jacobson, Nathan. Lie and Jordan triple systems, ... 149

Kodaira, Kunihiko. The eigenvalue problem for ordinary differential equations of the second order and Heisenberg's theory of S-matrices, ... 921

Lehner, Joseph. Divisibility properties of the Fourier coefficients of the modular invariant $j(\tau)$, ... 136

Lehner, Joseph. Further congruence properties of the Fourier coefficients of the modular invariant $j(\tau)$, ... 373

Lewis, D. C. Metric properties of differential equations, ... 294

Liao, S. D. Some theorems on the dimension of fibre spaces, . . 231

Loo, Ching-Tsün. Note on the properties of Fourier coefficients, . 269

Lorentz, G. G. A problem of plane measure, ... 417

McCoy, Neal H. Prime ideals in general rings, ... 823

Minakshisundaram, S. Notes on Fourier expansions III, . . . 60

Mitchell, Josephine. A summability theorem for double orthogonal series whose coefficients satisfy certain conditions, . . . 257

Nachbin, Leopoldo. A characterization of the normed vector ordered spaces of continuous function over a compact space, . . . 701

Nakayama, Tadasi. Semilinear normal basis for quasifields, . . 241

Nassif, M. On the mode of increase of the product of basic sets of polynomials, ... 40

Nehari, Zeev. On the accessory parameters of a Fuchsian differential equation, ... 24

Nehari, Zeev. The radius of univalence of an analytic function, . 845

Neumann, B. H. On ordered groups, ... 1

Neumann, Hanna. Generalized free products with amalgamated subgroups, ... 491

PAGE

Peck, L. G. On the approximation of irrational numbers by the convergents of their Diophantine equations in algebraic number fields, ... 387

Putnam, C. R. On the spectra of certain boundary value problems, 109

Putnam, C. R. The cluster spectra of bounded potentials, . . 612

Reid, William T. Expansion methods for the isoperimetric problem of Bolza in non-parametric form, ... 946

Rich, R. P. Completely simple ideals of a semigroup, . . . 883

Rosenthall, E. Multiplicative diophantine equations in quaternions, 791

Sard, Arthur. Best approximate integration formulas; best approximation formulas, ... 80

Shapiro, George. On the non-vanishing at $s=1$ of certain Dirichlet series, ... 621

Schafer, R. D. Structure of genetic algebras, ... 121

Thron, W. J. Twin convergence regions for continued fractions $b_0+K(1/b_n)$, II, ... 112

Vermes, P. Series to series transformations and analytic continuation by matrix methods, ... 541

Weyl, Hermann. Almost periodic invariant vector sets in a metric vector space, ... 178

Whyburn, G. T. Continuous decompositions, ... 218

Wintner, Aurel. On a decomposition into singularities of theta-functions of fractional index, ... 105

Wintner, Aurel. See Hartman and Wintner, page 206.

Wintner, Aurel. See Hartman and Wintner, page 214.

Wintner, Aurel. On the classical existence theorem of linear differential equations, ... 331

Wintner, Aurel. On linear repulsive forces, ... 362

Wintner, Aurel. See Hartman and Wintner, page 367.

Wintner, Aurel. A priori Laplace transformations of linear differential equations, ... 587

Wintner, Aurel. On almost free linear motions, ... 595

Wintner, Aurel. On the smallness of isolated eigenfunctions, . . 603

Wintner, Aurel. See Hartman and Wintner, page 627.

Wintner, Aurel. See Hartman and Wintner, page 650.

Wintner, Aurel. On linear asymptotic equilibria, ... 853

Wintner, Aurel. See Hartman and Wintner, page 859.

Wintner, Aurel. See Hartman and Wintner, page 865.

Wintner, Aurel. See Coddington and Wintner, page 886.

Yood, Bertram. Additive groups and linear manifolds of transformations between Banach spaces, ... 663

Young, Gail S., Jr. On 1-regular convergence of sequences of 2-manifolds, ... 339

CALCUTTA CENTRAL LIBRARY UNIVERSITY

ON ORDERED GROUPS.*

By B. H. NEUMANN.

Certain axiomatic questions in geometry lead to the study of ordered division rings (cf. Hilbert [8]), and these in turn to the study of ordered groups: I only mention (without proof) that every ordered group can be embedded in (the multiplicative group of) an ordered division ring.

F. W. Levi [9], [10] has given necessary conditions (not sufficient), and also sufficient conditions (not necessary), for a group to be capable of being ordered. In the first section of this paper the necessary conditions are generalised, a typical result being: If two elements of an ordered group are not permutable with each other, then none of their powers are. A similar result is also proved for certain higher commutators. Just how far one can proceed in this direction remains an open question: Some light is thrown on it by an example.

The sufficient conditions of Levi (*loc. cit.*) can also be generalised. In the second section we derive a very general sufficient criterion (which is also necessary, but only trivially so). An order is actually constructed in a group when the criterion applies; if the group possesses an ordered factor group, then its order can be utilised in the construction.

The general criterion is an unwieldy weapon. It can be specialised in various ways (**3**). From one of these specialisations one sees that all free groups can be ordered. (This result has also been obtained by G. Birkhoff and independently A. Tarski; cf. G. Birkhoff [3].) More generally we show that the order of any ordered group can be refined to an order of a free group of which the given group is a homomorphic image. Some special constructions of new ordered groups from given ordered groups are also given.

In the fourth and final section these constructive methods are used to construct an ordered group which coincides with its commutator group. Such an example throws some light on the limitations of the various criteria and other results; it may also be of interest in itself, and is given in some detail.

1. Necessary conditions for ordered groups. We call a group G an 0-group if it can be fully ordered, i. e., if a transitive binary relation $a < b$ can be defined in G, such that of the three alternatives $a < b$, $a = b$, $b < a$

---

* Received February 13, 1947.

one and only one takes place, and $a < b$ implies $at < bt$ and $ta < tb$ for all a, b, t in G. If G is an 0-group, and an order relation has been chosen for G, we call G simply an ordered group.[1] The group consisting of the unit element only is an "improper" ordered group (with void order).

We write G multiplicatively, denote the unit element by 1, and the order relation by $<$, even when dealing with several groups simultaneously: different order relations will be distinguished by the context. We also use the commutator notation of P. Hall [7]; thus

$$[x, y] = x^{-1}y^{-1}xy,$$
$$[x, y, z] = [[x, y], z],$$

and so on, with the corresponding notation for subgroups.

We call a group "locally infinite" if every element $\neq 1$ in it is of infinite order.[2] The group which consists of the unit element only is "improperly" locally infinite.

Levi [9] shows that an 0-group is locally infinite, and more generally that the equation $x^m = a$ for $a \varepsilon G$, m a natural number, has at most one solution x. One can show more generally:

1.1 LEMMA. *If a, b are elements of an 0-group G, and $[a^m, b] = 1$ for any integer $m \neq 0$, then $[a, b] = 1$.*

Proof. Assume that $[a, b] \neq 1$, and that $[a, b] > 1$ in some order of G. Now

$$[a^m, b] \equiv \prod_{\mu=m-1}^{\mu=0} (a^{-\mu}[a, b]a^{\mu})$$

and

$$[a^{-m}, b] \equiv \prod_{\mu=-m}^{\mu=-1} (a^{-\mu}[a, b]^{-1}a^{\mu})$$

for all $m > 0$. Hence $[a^m, b]$ is a product of conjugates of $[a, b]$; each of these is > 1, and therefore $[a^m, b] > 1$. Also $[a^{-m}, b]$ is a product of conjugates of $[a, b]^{-1}$; each of these is < 1, and therefore $[a^{-m}, b] < 1$. Hence if $[a, b] \neq 1$, then $[a^m, b] \neq 1$, and the lemma follows.

1.2 COROLLARY. *If two elements of an 0-group are not permutable with each other, then none of their powers ($\neq 1$) are permutable with each other.*

---

[1] This is, of course, a special case of the *o*-groups of Everett and Ulam [4] and *l*-groups of G. Birkhoff [2].

[2] A group has a property locally if all its subgroups of finite rank (generated by a finite number of elements) have the property.

We can further extend 1.1 by establishing the following necessary condition for 0-groups.

1.3 Lemma. *If a, b are elements of an 0-group G, and $[a^m, b, a] = 1$ for any integer $m \neq 0$, then $[a, b, a] = 1$.*

Proof. Let m again be a *positive* integer. Then we expand

$$[a^m, b, a] \equiv [\prod_{\mu=m-1}^{\mu=0} (a^{-\mu}[a, b]a^{\mu}), a]$$

$$\equiv \prod_{\mu=m-1}^{\mu=0} (t_\mu^{-1}[a^{-\mu}[a, b]a^{\mu}, a]t_\mu),$$

where t_μ are certain part products of $\prod_{\mu=m-1}^{\mu=0} (a^{-\mu}[a, b]a^{\mu})$. Then

$$[a^m, b, a] \equiv \prod_{\mu=m-1}^{\mu=0} ((a^\mu t_\mu)^{-1}[a, b, a](a^\mu t_\mu));$$

hence again $[a^m, b, a]$ is a product of conjugates of $[a, b, a]$, and therefore is $\gtrless 1$ if $[a, b, a] \gtrless 1$. Similarly

$$[a^{-m}, b, a] \equiv [\prod_{\mu=-m}^{\mu=-1} (a^{-\mu}[a, b]^{-1}a^{\mu}), a]$$

$$\equiv \prod_{\mu=-m}^{\mu=-1} (t'^{-1}_\mu[a^{-\mu}[a, b]^{-1}a^{\mu}, a]t'_\mu)$$

$$\equiv \prod_{\mu=-m}^{\mu=-1} ((a^\mu t'_\mu)^{-1}[[a, b]^{-1}, a](a^\mu t'_\mu))$$

$$\equiv \prod_{\mu=-m}^{\mu=-1} (([a, b]^{-1}a^\mu t'_\mu)^{-1}[a, b, a]^{-1}([a, b]^{-1}a^\mu t'_\mu)).$$

Thus $[a^{-m}, b, a]$ is a product of conjugates of $[a, b, a]^{-1}$ and therefore is $\lessgtr 1$ if $[a, b, a] \gtrless 1$. The lemma follows.

1.4 Corollary. *If a, b are elements of an 0-group G and their commutator $[a, b]$ is not permutable with a, then no commutator $[a^m, b]$ of a power $(\neq 1)$ of a and b is permutable with any power $(\neq 1)$ of a.* Or: *if $[a, b, a] \neq 1$, then $[a^m, b, a^n] \neq 1$ for any $m \neq 0$, $n \neq 0$.*

This follows by applying Lemmas 1.1 and 1.3.

One will naturally look for further generalisations of these necessary conditions for 0-groups. Two directions suggest themselves: One, to decide whether $[a^m, b, a, a] = 1$ $(m \neq 0)$ entails $[a, b, a, a] = 1$; the other, to decide whether $[a^m, b, c] = 1$ $(m \neq 0)$ entails $[a, b, c] = 1$. The first of these questions I can not answer; the second is answered, negatively, by the following example.

1.5 *Example.* Let the group H be generated by elements

$$\cdots, b_{-1}, b_0, b_1, b_2, \cdots$$
$$\cdots, c_{-1}, c_0, c_1, c_2, \cdots$$

with the defining relations

1.51 $[b_{\mu+1}, b_\mu] = c_\mu,$ $\qquad \mu = 0, \pm 1, \pm 2, \cdots$

1.52 $[b_{\mu+\nu}, b_\mu] = 1,$ $\qquad \mu = 0, \pm 1, \pm 2, \cdots$ $\nu = 2, 3, \cdots$

1.53 $[b_\mu, c_\nu] = 1,$ $\qquad \mu, \nu = 0, \pm 1, \pm 2, \cdots$

1.54 $[c_\mu, c_\nu] = 1,$ $\qquad \mu, \nu = 0, \pm 1, \pm 2, \cdots.$

We define an order in H such that

1.55 $1 < \cdots \ll c_{-1} \ll c_0 \ll c_1 \ll \cdots \ll b_{-1} \ll b_0 \ll b_1 \ll \cdots$

where $x \ll y$ means that all powers of x lie between y^{-1} and y. Thus any element of H is > 1 if the highest-suffix b_μ in it appears with positive exponent, or if there is no b_μ in it, and the highest-suffix c_μ in it appears with positive exponent. One can satisfy oneself without difficulty that in this way H does become an ordered group. The mapping

$$b_\mu \rightarrow b_{\mu+1}, \qquad c_\mu \rightarrow c_{\mu+1}$$

clearly defines an automorphism of H *qua* group, and this automorphism leaves the order of H invariant. We now define G by adjoining this automorphism to H, i. e., we form $G = \{H, a\}$ with the relations

1.56 $a^{-1} b_\mu a = b_{\mu+1};$ $\quad a^{-1} c_\mu a = c_{\mu+1};$ $\qquad \mu = 0, \pm 1, \pm 2, \cdots$

and we order G by making $a > 1$ and $a \gg h$ for all $h \varepsilon H$. It is again easy to see that G thus becomes an ordered group.

Now in G

$$[a, b_0, b_0] = [a^{-1} b_0^{-1} a b_0, b_0] = [b_1^{-1} b_0, b_0] = [b_1^{-1}, b_0] = c_0^{-1} < 1,$$

but when $m > 1$

$$[a^m, b_0, b_0] = [a^{-m} b_0^{-1} a^m b_0, b_0] = [b_m^{-1} b_0, b_0] = 1.$$

This example, therefore, proves

1.6 Lemma. *In an 0-group,* $[a, b, b] \neq 1$ *is compatible with* $[a^m, b, b] = 1$ *for* $m > 1$.

and thus *a fortiori*

1.7 COROLLARY. *In an 0-group,* $[a,b,c] \neq 1$ *is compatible with* $[a^m,b,c] = 1$ *for* $m > 1$.

2. Sufficient conditions for ordered groups. We use the following criterion for 0-groups, adapted from one given by Levi [9].

2.1 LEMMA. *The group* G *is an 0-group if (and only if) it contains two subsets* s^+ *and* s^- *such that*

2.11 $$s^+ \cup s^- = G - \{1\},$$

i. e., every element of G*, except the unit element, lies in* s^+ *or in* s^-;

2.12 $$s^+ \cdot s^+ \subset s^+ \text{ and } s^- \cdot s^- \subset s^-,$$

i. e., s^+ *and* s^- *are semi-groups;*

2.13 $$t^{-1}s^+t \subset s^+ \text{ for all } t \in G,$$

i. e., s^+ *(and therefore also* s^-*) is self-conjugate in* G.

Proof. Let G possess two such subsets. Then, as they are semi-groups but do not contain the unit element, neither of them contains an element simultaneously with its inverse. But between them they contain all elements of G except 1. Hence of a pair of inverse elements one always belongs to s^+ and the other to s^-. Now we define an order relation in G by

2.14 $$a < b \text{ if and only if } a^{-1}b \in s^+.$$

Then if $a < b$, $b^{-1}a \in s^-$, and therefore $b \nless a$; also $b \neq a$, as the unit element is not in s^+. Hence of the three alternatives $a < b$, $a = b$, $b < a$, not more than one takes place. But one of them does take place, as of the two inverse elements $a^{-1}b$, $b^{-1}a$ one lies in s^+, unless $a^{-1}b = 1$. Also if $a < b$ and $b < c$, then $a^{-1}b \in s^+$, $b^{-1}c \in s^+$; hence $a^{-1}c \in s^+$, by 2.12, and $a < c$; which shows transitivity of the order relation. Finally if $a^{-1}b \in s^+$ then also $(at)^{-1}bt \in s^+$ because of 2.13, and $(ta)^{-1}tb \in s^+$ trivially; i. e. if $a < b$ then $at < bt$ and $ta < tb$. Hence G is ordered.

The converse is also true; for if G is an 0-group, we choose an order of G and then denote by s^+ the set of all elements > 1, by s^- the set of all elements < 1. Then 2.11-13 are easily checked.

We shall also use the following sufficient criterion, due to Levi [9] (*q. v.* for a proof):

2.2 THEOREM. *A locally infinite abelian group is an 0-group.*

The most general sufficient criterion that we derive is the following.

2.3 THEOREM. *Let the group G possess a set of subgroups linearly ordered by inclusion:*

2.30
$$(G \supset) \cdots \supset H_\alpha \supset H_{\alpha'} \supset \cdots \supset \{1\}$$

(not necessarily all different) with the following properties:

2.31 *Each H_α is self-conjugate in G, i.e., 2.30 is a generalised normal series of G.*

2.32 *Each H_α except $\{1\}$ has an immediate successor $H_{\alpha'}$ in the series 2.30.*[3]

2.33 *For all terms of the series,*

$$[G, H_\alpha] \subset H_{\alpha'},$$

i.e., 2.30 is a generalised central series of any one of its terms.

2.34 *If 2.30 has a first term H_1, then G/H_1 is an 0-group.*

2.35 *$H_\alpha/H_{\alpha'}$ is (properly or improperly) locally infinite.*

2.36 *To every element $g \in H_1$ if 2.30 has a first term H_1, or to every element $g \in G$ if 2.30 has no first term, there is a minimal H_α containing it, i.e., $g = 1$ or there is an H_α with $g \in H_\alpha - H_{\alpha'}$. Then G is an 0-group.*

Proof. By 2.33 and 2.35 each $H_\alpha/H_{\alpha'}$ is abelian and locally infinite, hence, by 2.2, an 0-group. In each $H_\alpha/H_{\alpha'}$ we choose[4] an order. We denote by $s^+{}_\alpha H_{\alpha'}$ ($s^-{}_\alpha H_{\alpha'}$) the set of all elements of H_α which are greater (smaller) than the unit element (mod $H_{\alpha'}$) in this order. If 2.30 possesses a first term H_1, we also choose an order of G/H_1, and define $s^+{}_0 H_1$ ($s^-{}_0 H_1$) accordingly. Finally we introduce the union s^+ (s^-) of all $s^+{}_\alpha H_{\alpha'}$ ($s^-{}_\alpha H_{\alpha'}$)

2.41
$$s^+ = \bigcup_\alpha s^+{}_\alpha H_{\alpha'}, \qquad s^- = \bigcup_\alpha s^-{}_\alpha H_{\alpha'}.$$

Now let $g \neq 1$ be an element of G; then (by 2.36) either there is an H_α such that $g \in H_\alpha - H_{\alpha'}$, or 2.30 has a first term H_1 and $g \in G - H_1$. Hence g is greater or smaller than the unit element (mod $H_{\alpha'}$ or mod H_1) in H_α or G: hence $g \in s^+{}_\alpha H_{\alpha'} \cup s^-{}_\alpha H_{\alpha'}$ (or $s^+{}_0 H_1 \cup s^-{}_0 H_1$). In any case $g \in s^+ \cup s^-$, and

2.42
$$s^+ \cup s^- = G - \{1\}.$$

---

[3] Note that the series 2.30 need not be well-ordered; its order type is finite or made up, additively, from order types ω and $^*\omega + \omega$, with a finite tail. (For the notation *cf.* Fraenkel [5]).

[4] The axiom of choice is used. 2.2 requires well-order for its proof.

Let now g and h be two elements in s^+. Let $g \varepsilon H_\alpha - H_{\alpha'}$, $h \varepsilon H_\beta - H_{\beta'}$, and $H_\beta \subset H_\alpha$, say. Now if $H_\beta \neq H_\alpha$, i. e., $H_\beta \subset H_{\alpha'}$, then gh is congruent to g (mod $H_{\alpha'}$); then $gh \varepsilon s^+{}_\alpha H_{\alpha'}$ and $gh \varepsilon s^+$. If $H_\beta = H_\alpha$, then g and h are both in the same $s^+{}_\alpha H_{\alpha'}$, hence their product is. Hence again $gh \varepsilon s^+$. Correspondingly if g or h lies outside H_1. In any case

2. 43 $$s^+ \cdot s^+ \subset s^+.$$

Similarly one proves

$$s^- \cdot s^- \subset s^-.$$

Finally let $g \varepsilon s^+$, let us say $g \varepsilon s^+{}_\alpha H_{\alpha'}$; and $t \varepsilon G$ arbitrary. Then [5]

$$t^{-1}gt = g \cdot [g, t] \varepsilon g \cdot [H_\alpha, G] \subset g \cdot H_{\alpha'}.$$

Thus

$$t^{-1}gt \cdot H_{\alpha'} = g \cdot H_{\alpha'},$$

and so

$$t^{-1}gt \varepsilon s^+{}_\alpha H_{\alpha'}.$$

If, on the other hand, 2. 30 has a first term and $g \varepsilon s^+{}_0 H_1$ then also

$$t^{-1}gt \varepsilon s^+{}_0 H_1;$$

for $s^+{}_0$ is self-conjugate in G/H_1, and H_1 is self-conjugate in G; hence $s^+{}_0 H_1$ is self-conjugate in G. Thus we see that

2. 44 $$t^{-1}s^+t \subset s^+ \quad \text{for all } t \varepsilon G.$$

Combining 2. 42-2. 44 and 2. 1, we see that G is an 0-group and the proof of the theorem is complete.

If G is an ordered group, H a self-conjugate subgroup of G, and $K = G/H$ is ordered so that $aH \leqq bH$ in the order of K whenever $a < b$ in the order of G, then we call the order of G a "refinement" of the order of K. Then the proof of Theorems 2. 3 also shows:

2. 5 COROLLARY. *If under the conditions of the criterion* 2. 3 *the series* 2. 30 *has a first term* H_1, *then any order of* G/H_1 *can be refined to an order of* G.

---

[5] This is the only step in the proof which fully uses 2. 33. One easily sees that 2. 33 and 2. 35 can be replaced by the following condition (which is not, however, equivalent to 2. 33, 2. 35)

2. 33′ Each $H_\alpha/H_{\alpha'}$ is an 0-group, and in particular possesses an order which admits all inner automorphisms of G.

Then for the purposes of the proof such an order has to be chosen in each $H_\alpha/H_{\alpha'}$. This modification of the theorem generalizes another criterion due to Levi [9].

3. Special methods for constructing ordered groups. Theorem 2. 3 is somewhat unwieldy. Some special cases may, however, be of interest.

Let G be a group and denote by ${}^nG$ the terms of its lower central series:

$$ {}^0G = G, \qquad {}^{n+1}G = [G, {}^nG]. $$

Further denote by $Z_n(G)$ the terms of its upper central series:

$$ Z_0(G) = \{1\}, \qquad Z_{n+1}(G)/Z_n(G) \text{ the centre of } G/Z_n(G). $$

3. 1 THEOREM. *If G is such that*[6]

3. 11 $$\bigcap_n {}^nG = \{1\};$$

3. 12 *${}^nG/{}^{n+1}G$ is locally infinite for $n = 0, 1, 2, \cdots$; then G is an 0-group.*

3. 2 THEOREM. *If G is such that*[7]

3. 21 $$\bigcup_n Z_n(G) = G;$$

3. 22 *$Z_{n+1}(G)/Z_n(G)$ is locally infinite for $n = 0, 1, 2, \cdots$; then G is an 0-group.*

Both these theorems are easy consequences of 2. 3. The set of groups H_α, consisting of the terms of the lower or upper central series, is here finite or of order type ω (3. 1) or ${}^*\omega$ (3. 2), so that 2. 32 is satisfied. The definition of the central series assures 2. 31, 2. 33. Assumptions 3. 11 and 3. 21 entail 2. 36, 3. 12 and 3. 22 are simply 2. 35. 2. 34 is also satisfied, as a consequence of 2. 2. Hence 2. 3 applies.

3. 3 COROLLARY. (*Cf.* Birkhoff [3]) *All free groups (of finite or infinite rank) are 0-groups.*

3. 4 THEOREM. *If the ordered group G is represented as a factor group of a free group F with respect to a relation group R,*

3. 41 $$G \cong F/R,$$

then the order of G can be refined to an order of F.

Proof. We use the intersection of R with the terms ${}^nF$ of the lower central series of F, putting

3. 42 $$R_0 = (R \cap {}^0F =) R, \qquad R_n = R \cap {}^nF, \qquad n = 0, 1, 2, \cdots.$$

Now as R is self-conjugate in F, R_n is also self-conjugate in F, and

---

[6] 3. 11 means that G is an N-group in the terminology of Baer [1].

[7] 3. 21 means that G is a Z-group in the terminology of Baer [1].

3.43 $$[F, R_n] \subset R_n \subset R.$$

Also

$$[F, R_n] \subset [F, {}^nF] = {}^{n+1}F.$$

Hence

3.44 $$[F, R_n] \subset R_{n+1}.$$

Let $a \varepsilon F$ have a (proper) power in R_{n+1}

$$a^k \varepsilon R_{n+1}, \qquad k \neq 0.$$

Then $a^k \varepsilon {}^{n+1}F$; but as $F/{}^{n+1}F$ is locally infinite, $a \varepsilon {}^{n+1}F$. Also $a^k \varepsilon R$; but as F/R is an 0-group and therefore locally infinite, also $a \varepsilon R$. Hence

$$a \varepsilon {}^{n+1}F \cap R = R_{n+1}.$$

This means that F/R_{n+1} is locally infinite, and therefore *a fortiori* R_n/R_{n+1} is locally infinite.

Finally

3.45 $$\bigcap_n R_n \subset \bigcap_n {}^nF = \{1\}.$$

Hence every element $r \varepsilon R$, $r \neq 1$ is in a smallest R_n,

$$r \varepsilon R_n - R_{n+1}.$$

The theorem follows now simply from Theorem 2.3 and Corollary 2.5.

We now give some fairly obvious results which can be used for constructing new ordered groups from given ordered groups. Detailed proofs are omitted.

3.5 The (complete[8]) direct product of any set of 0-groups is an 0-group.

For we can well-order the direct factors, and chose an order in each. The direct product is then simply ordered by the convention that an element is $\gtrless 1$ according as the first component (in the well-order of the direct factors) $\neq 1$[9] of the element is $\gtrless 1$ in the chosen order of the factor.

The following construction dispenses with well-order.

3.6 Given an ordered set of 0-groups, their *restricted* direct product[10] can be so ordered that the direct factors appear in the given set order.

For we can choose an order in each direct factor. The restricted direct product is then simply ordered by the convention that an element is $\gtrless 1$ according as the last component $\neq 1$ (in the order of the set of factors) of the element is $\gtrless 1$ in the chosen order of the factor.

---

[8] I. e. without restriction upon the number (or cardinal) of components $\neq 1$ of an element.

[9] 1 stands for the unit element of all the groups that occur; similarly $\lesseqgtr$ applies to the order chosen in the factors as well as to that under construction in the product.

[10] I. e., that in which every element has only a finite number of components $\neq 1$.

3. 5 and 3. 6 are in fact only special cases of known results. *Cf.* Hahn [6].

As an application of 3. 6 we give, in some detail, the following construction, which will be used in the next section.

3. 7 Starting from an ordered group B we form the restricted direct product of an ordered set of type $^*\omega + \omega$ of factors B_n, $n = 0, \pm 1, \pm 2 \cdots$ each isomorphic to B:

$$B^* = \cdots \times B_{-1} \times B_0 \times B_1 \times B_2 \times \cdots .$$

The elements are of the form

3. 71 $$b^* = \cdots \times b_{-1,-1} \times b_{0,0} \times b_{1,1} \times b_{2,2} \times \cdots ,$$

(where the first suffix distinguishes the direct factor in which the component lies, the second suffix the element of B which appears as this component); only a finite number of the components $b_{n,n}$ are different from the unit element, and the last one of these determines whether $b^* \gtrless 1$. Now B^* possesses an obvious automorphism (relating to its order as well as to the group operation), viz. that mapping each component B_n on its successor B_{n+1}. We denote this automorphism by a and extend B^* by means of it; i. e. we form all the products

3. 72 $$g = a^{\alpha} b^*$$

with the transformation rule

3. 73 $$a^{-1} b^* a = \cdots \times b_{-1,-2} \times b_{0,-1} \times b_{1,0} \times b_{2,1} \times \cdots$$

(where b^* is given by 3. 71). The elements 3. 72 form a group G, which we order first according to the power of a in it, then according to b^*. Thus

3. 74 $$g \gtrless 1 \text{ if } \alpha \gtrless 0, \text{ or if } \alpha = 0 \text{ and } b^* \gtrless 1.$$

If B is given by a system of generators $b, b', \cdots$ and defining relations $r(b, b', \cdots) = r'(b, b', \cdots) = \cdots = 1$, then we can give G in the same manner

3. 75 $$\begin{aligned} G = \{a, b, b', \cdots ;\ & r(b, b', \cdots) = r'(b, b', \cdots) = \cdots = 1, \\ & [a^{-\lambda} b a^{\lambda}, a^{-\mu} b a^{\mu}] = [a^{-\lambda} b a^{\lambda}, a^{-\mu} b' a^{\mu}] = \cdots = 1, \\ & (\lambda \neq \mu)\}. \end{aligned}$$

The commutator relations are formed for all pairs of generators $b, b', \cdots$

of B, but may then be restricted to $\lambda = 0$, $\mu > 0$. If G is defined by 3.75, the order in G can be described without reference to B^*, solely in terms of a, B. To this end we represent an element of G in the form

3.76 $$g = a^{\alpha} \prod_{i=1}^{i=m} a^{-\mu_i} b_i a^{\mu_i}$$

where $\mu_1 > \mu_2 > \cdots > \mu_m$, all $b_i \neq 1$.[11] This is possible because transforms of elements of B by different powers of a are permutable with each other; α is the sum of exponents with which a appears in g.[12] Now the order is defined in G by

3.77 $$g \gtrless 1 \text{ if } \alpha \gtrless 0 \text{ or } \alpha = 0 \text{ and } b_1 \gtrless 1.$$

This construction can be extended by replacing the powers of a by the elements of an arbitrary ordered group A.

3.8 Let the ordered groups A and B be given by

3.81 $$A = \{a, a', \cdots; q(a, a', \cdots) = q'(a, a', \cdots) = \cdots = 1\},$$

3.82 $$B = \{b, b', \cdots; r(b, b', \cdots) = r'(b, b', \cdots) = \cdots = 1\}.$$

Then we form the group

3.83 $$G = \{a, a', \cdots, b, b', \cdots; \quad 3.84\text{—}.86\}$$

where the relations are

3.84 $$q(a, a', \cdots) = q'(a, a', \cdots) = \cdots = 1,$$

3.85 $$r(b, b', \cdots) = r'(b, b', \cdots) = \cdots = 1,$$

3.86 $$[b, a_1^{-1} b a_1] = [b, a_1^{-1} b' a_1] = \cdots = [b, a_2^{-1} b a_2] = \cdots = 1.$$

In the commutator relations 3.86 $a_1, a_2, \cdots$ range over all elements > 1 of A, and the elements of B involved range over all pairs of generators $b, b', \cdots$. An element $g \varepsilon G$ can be represented in the form

3.87 $$g = a_0 \prod_{\mu=1}^{\mu=m} (a_\mu^{-1} b_\mu a_\mu)$$

with $a_1 > a_2 > \cdots > a_m$ in A, and with all $b_\mu \neq 1$. This is possible because transforms of elements of B by different elements of A are permutable with each other. Then G is ordered by the convention

---

[11] If $m = 0$, the product is void.

[12] Easily seen to be an invariant as long as a and elements of B are chosen as the generators of G.

3.88 $$g \gtrless 1 \text{ if } a_0 \gtrless 1 \text{ or } a_0 = 1 \text{ and } b_1 \gtrless 1.$$

The proof that in this way G becomes an ordered group is omitted.

Finally we mention, without proof, a construction principle due to Steinitz [11].

3.9 Let a system Σ of ordered groups G_α be given with the property that to any two groups G_α, G_β in Σ there is a group G_γ in Σ which contains both G_α and G_β as subgroups[13] and continues the order of both. Then there is an ordered group G containing as subgroups all the groups in Σ, each with its order, and generated by them.

4. A perfect ordered group. Levi [10] shows that if an 0-group G is finitely generated then it is different from its commutator group G'. To show that the finiteness of the number of generators can not be dispensed with; to illustrate the limitations to our various sufficient criteria for 0-groups (2.3, 3.1, 3.2); and to demonstrate the application of our various constructive principles: we now construct an ordered group which is perfect,[14] i.e. coincides with its commutator group.

Starting from an infinite cycle

4.01 $$H_1 = \{b_1\}$$

we first define a series H_n of groups by repeated application of 3.7.

4.02 $$H_n = \{b_1, b_2, \cdots, b_n;\ [b_r, b_q^{-\lambda} b_s b_q^{\lambda}] = 1 \ \ (q > r, s;\ \lambda \neq 0\text{[15]})\}.$$

Clearly H_n is obtained from H_{n-1} by adding the generator b_n and relations which entail that two elements of H_{n-1} transformed by different powers of b_n are permutable. The method of 3.7 can also be used to order H_n, when it can be seen that the order of H_n continues that of H_{n-1}; but we do not require the order at this stage.

We now consider the elements

4.03 $$c_\nu = [b_\nu, b_n] = b_\nu^{-1} \cdot b_n^{-1} b_\nu b_n, \qquad \nu = 1, 2, \cdots, n-1,$$

and denote by K_{n-1} the subgroup of H_n generated by these elements. We proceed to show that

4.04 $$K_{n-1} \cong H_{n-1};$$

more particularly, the mapping

---

[13] A group G_γ may contain different subgroups isomorphic and similarly ordered to G_α; but G_α must *be* one of the subgroups of G_γ.

[14] We use "perfect" in its group-theoretical sense.

[15] Here as later λ may be restricted to positive values.

$$c_\nu \leftrightarrow b_\nu, \qquad \nu = 1, 2, \cdots, n-1,$$

defines an isomorphism between K_{n-1} and H_{n-1}. To see this we form any word $w(b_1, b_2, \cdots, b_{n-1})$ in the generators of H_{n-1}. Then

$$4.05 \quad w(c_1, c_2, \cdots, c_{n-1}) = w(b_1{}^{-1}, b_2{}^{-1}, \cdots, b_{n-1}{}^{-1}) \cdot b_n{}^{-1} w(b_1, b_2, \cdots, b_{n-1}) b_n;$$

for the expressions in $c_\nu{}^{-1}$ permute with those in $b_n{}^{-1} b_\nu b_n$. The two factors on the right-hand side of 4.05 lie in different components (viz. H_{n-1} and $b_n{}^{-1} H_{n-1} b_n$) of a direct product. Hence the left-hand side of 4.05 can equal the unit element only if both factors on the right-hand side do. Thus

$$w(c_1, c_2, \cdots, c_{n-1}) = 1$$

entails

$$w(b_1, b_2, \cdots, b_{n-1}) = 1,$$

and H_{n-1} is a homomorphic image of K_{n-1} under the mapping $c_\nu \rightarrow b_\nu$.

Conversely let $w(b_1, b_2, \cdots, b_{n-1}) = 1$. Then w is a product of conjugates (in H_{n-1}) of the left-hand sides of the defining relations for $b_1, b_2, \cdots, b_{n-1}$. Now these defining relations express the permutability of transforms by different powers of b_q, of any two elements expressible in terms of $b_1, b_2, \cdots, b_{q-1}$. From these relations then follows also the permutability of transforms by different powers of $b_q{}^{-1}$, of any two elements expressible in terms of $b_1{}^{-1}, b_2{}^{-1}, \cdots, b_{q-1}{}^{-1}$. Thus if

$$w(b_1, b_2, \cdots, b_{n-1}) = 1$$

is a relation connecting the generators, then

$$w(b_1{}^{-1}, b_2{}^{-1}, \cdots, b_{n-1}{}^{-1}) = 1$$

is a relation connecting their inverses. Hence in this case the whole right-hand side of 4.05 equals the unit element, and

$$w(c_1, c_2, \cdots, c_{n-1}) = 1$$

follows from

$$w(b_1, b_2, \cdots, b_{n-1}) = 1.$$

The mapping $c_\nu \leftrightarrow b_\nu$ generates, therefore, an isomorphism between K_{n-1} and H_{n-1}.[16]

---

[16] Note that K_{n-1} does not, in general, contain all the elements $g_{n-1}{}^{-1} b_n{}^{-1} g_{n-1} b_n$. Thus it contains

$$c_1 c_2 = b_1{}^{-1} b_2{}^{-1} \cdot b_n{}^{-1} b_1 b_2 b_n,$$

but not

$$(b_1 b_2)^{-1} \cdot b_n{}^{-1} (b_1 b_2) b_n.$$

The intrinsic reason for the isomorphism of K_{n-1} and H_{n-1} is interesting, but beyond the scope of this paper.

Now similarly H_{n-1} contains in its commutator group a subgroup isomorphic to H_{n-2}; hence K_{n-1} contains in its commutator group a subgroup[17] L_{n-2} isomorphic to H_{n-2}; and so it goes on. The idea of the construction is now to consider a sequence of groups $\cdots, L, K$, rather than that of the groups H; in this way we ensure that each term of the sequence lies in the commutator group of its successor.[18] To do this we define groups G_n each isomorphic to H_n; but such that an isomorphism from G_n to H_n maps G_{n-1} on K_{n-1}, not on H_{n-1}. We define

4.06 $\quad G_n = \{a_{11}, a_{21}, a_{22}, a_{31}, a_{32}, a_{33}, \cdots, a_{n1}, \cdots, a_{nn};\ 4.07,\ .08\}$
with the relations

4.07 $\quad [a_{pr}, a_{pq}^{-\lambda} a_{ps} a_{pq}^{\lambda}] = 1$ for $n \geq p \geq q > r, s$; $\lambda \neq 0$;

4.08 $\quad [a_{pq}, a_{pp}] = a_{p-1,q}$ for $n \geq p > q$.

It is seen from 4.08 that G_n can be generated by $a_{n1}, a_{n2}, \cdots, a_{nn}$; and those relations 4.07 for which $p = n$ are the same as the defining relations of H_n in 4.02. Therefore the mapping

$$b_1 \to a_{n1},\ b_2 \to a_{n2}, \cdots, b_n \to a_{nn}$$

generates a homomorphism of H_n onto G_n.

To show that this homomorphism is an isomorphism we prove that all the relations 4.07 follow already from those for which $p = n$, together with 4.08. The set of relations 4.07 for which p has acertain fixed value, $p = m$, say, will be denoted by 4.07_m for short; similarly we denote by 4.08_m those relations 4.08 for which $p = m$. We show that 4.07_{m-1} follow from 4.07_m and 4.08_m; then 4.07_p for $p = 1, 2, \cdots, n-1$ follow from 4.07_n together with 4.08.

Consider any word w formed of $m-1$ generators, $w(x_1, x_2, \cdots, x_{m-1})$. Then, using 4.08,

$$\begin{aligned} w(a_{m-1,1}, a_{m-1,2}, \cdots, a_{m-1,m-1}) &= w([a_{m1}, a_{mm}], [a_{m2}, a_{mm}], \cdots [a_{mm-1}, a_{mm}]) \\ &= w(a_{m1}^{-1} \cdot a_{mm}^{-1} a_{m1} a_{mm}, a_{m2}^{-1} \cdot a_{mm}^{-1} a_{m2} a_{mm}, \cdots, a_{mm-1}^{-1} \cdot a_{mm}^{-1} a_{mm-1} a_{mm}). \end{aligned}$$

By 4.07_m each $a_{mm}^{-1} a_{ms} a_{mm}$ permutes with each a_{mr}, for $r, s = 1, 2, \cdots, m-1$. Hence

---

[17] L_{n-2} is not a subgroup of H_{n-1}; hence its relation to H_{n-1} is not the same as that of K_{n-1} to H_n.

[18] If two sequences of groups are given

$$A_1 \subset A_2 \subset \cdots \text{ and } B_1 \subset B_2 \cdots$$

such that $A_1 \cong B_1$, $A_2 \cong B_2, \cdots$, and if A and B are the groups generated by these sequences by the Steinitz method (*cf.* 3.9), then A and B need not be isomorphic.

4.09 $$w(a_{m-1,1}, \cdots, a_{m-1,m-1}) = w(a_{m1}^{-1}, \cdots, a_{mm-1}^{-1}) \cdot a_{mm}^{-1} w(a_{m1}, \cdots, a_{mm-1}) a_{mm}$$

We apply this in particular to the left-hand side of 4.07$_{m-1}$, and obtain for $m - 1 \geq q > r, s;\ \lambda \neq 0$,

4.10 $$[a_{m-1,r}, a_{m-1,q}^{-\lambda} a_{m-1,s} a_{m-1,q}^{\lambda}]$$
$$= [a_{mr}^{-1}, a_{mq}^{\lambda} a_{ms}^{-1} a_{mq}^{-\lambda}] \cdot a_{mm}^{-1} [a_{mr}, a_{mq}^{-\lambda} a_{ms} a_{mq}^{\lambda}] a_{mm}.$$

Here both commutators on the right-hand side equal the unit element by 4.07$_m$, and 4.07$_{m-1}$ follows.

Hence all the relations in G_n follow from 4.07$_n$ and 4.08. The latter are only explicit definitions of the generators a_{pq}, $p < n$, one for each, and no relations between $a_{n1}, a_{n2}, \cdots, a_{nn}$ can follow from them. If we generate G_n by means of $a_{n1}, a_{n2}, \cdots, a_{nn}$ only, then 4.07$_n$ form a *complete* system of defining relations. Therefore the mapping generated by

4.11 $$a_{n1} \leftrightarrow b_1,\ a_{n2} \leftrightarrow b_2, \cdots, a_{nn} \leftrightarrow b_n$$

is an isomorphism between G_n and H_n. It is seen without difficulty that this isomorphism maps G_{n-1} on K_{n-1}; but the groups H_n and K_{n-1} are now no longer needed.

To define order in G_n we proceed as in 3.7; but in order to show that this order continues the order correspondingly defined for a subgroup $G_m \subset G_n$ $(m < n)$, we define the order simultaneously in the subgroup. To order G_m we form the chain of subgroups

4.12 $$G_{m1} = \{a_{m1}\},\ G_{m2} = \{a_{m1}, a_{m2}\}, \cdots, G_{mm} = \{a_{m1}, a_{m2}, \cdots, a_{mm}\}.$$

Then

4.13 $$G_{m1} \subset G_{m2} \subset \cdots \subset G_{mm} = G_m.$$

We proceed by induction. G_{m1} can be trivially ordered: $a_{m1}^{\lambda} \gtrless 1$ according as $\lambda \gtrless 0$. We assume that $G_{m,q-1}$ has been ordered already. Now let $g \neq 1$ be an element of $G_{m,q}$. Then g can be expressed in the form (cf. 3.7)

4.14 $$g = a_{mq}^{\lambda} \prod_i a_{mq}^{-\mu_i} g_i a_{mq}^{\mu_i},$$

where $\mu_1 > \mu_2 > \cdots$ and all $g_i \varepsilon G_{m,q-1}$.[19] Then we define

4.15 $$g \gtrless 1 \text{ if } \lambda \gtrless 0 \text{ or } \lambda = 0 \text{ and } g_1 \gtrless 1 \text{ (in } G_{m,q-1}).$$

It is easy to confirm the usual properties of this order relation, and we omit the proof.

---

[19] The product Π_i may consist of a single factor, or be absent.

To compare the order relations in G_m and G_{m-1} let $g \varepsilon G_{m-1}$ be expressed as a word

4.16 $$g = w(a_{m-1,1}, a_{m-1,2}, \cdots, a_{m-1,m-1}).$$

Then in G_m it can be expressed in the form 4.09, or preferably in the form

4.17 $$g = a_{mm}^{-1} w(a_{m1}, \cdots, a_{mm-1}) a_{mm} \cdot w(a_{m1}^{-1}, \cdots, a_{mm-1}^{-1}) .$$

This is of the form 4.14 with $q = m$, $\lambda = 0$, two factors in the product, $\mu_1 = 1$, $\mu_2 = 0$. Hence $g \gtreqless 1$ according as $w(a_{m1}, \cdots, a_{mm-1}) \gtreqless 1$. But it is clear that

$$w(a_{m1}, \cdots, a_{mm-1}) \gtreqless 1$$

according as

$$w(a_{m-1,1}, \cdots, a_{m-1,m-1}) \gtreqless 1;$$

for the first suffix m of the generators does not enter the definition 4.15 at all. Hence $g \gtreqless 1$ *qua* element of G_m according as $g \gtreqless 1$ *qua* element of G_{m-1}. By induction one then sees that the order of G_n coincides in G_m $(m < n)$ with the order of G_m.

We now have all the material together to construct the example, by applying Steinitz' method 3.9 to the series.

$$G_1, G_2, \cdots, G_n, \cdots.$$

4.2 *Example.* Let G_ω be the group generated by

$$a_{11}, a_{21}, a_{22}, a_{31}, a_{32}, a_{33}, \cdots, a_{n1}, \cdots, a_{nn}, \cdots$$

with the defining relations

4.21 $$[a_{pr}, a_{pq}^{-\lambda} a_{ps} a_{pq}^{\lambda}] = 1 \text{ for } p \geq q > r, s;\ \lambda \neq 0;$$

4.22 $$[a_{pq}, a_{pp}] = a_{p-1,q} \text{ for } p > q.$$

Relations 4.22 ensure that $G'_\omega = G_\omega$. Let G_ω be ordered by the definition 4.15 when the element $g \varepsilon G_\omega$ is expressed in the form 4.14. Then G_ω is a perfect ordered group.

As G_ω coincides with its commutator group, its lower central series is stillborn. So is its upper central series, for the center of G_ω is easily seen to be $\{1\}$. One can show even more:

4.3 Lemma. *Every element > 1 in G_ω has arbitrarily large conjugates: if $1 < g < h$ in G_ω, then there is an element $t \varepsilon G_\omega$ such that $t^{-1} g t > h$.*

Proof. Let m be such that g and h both lie in G_{m-1}. We express them in the next higher group G_m, using the representation 4.17, but abbreviating it to

4.31 $$g = a_{mm}^{-1} g_1 a_{mm} \cdot \bar{g}_1,$$

4.32 $$h = a_{mm}^{-1} h_1 a_{mm} \cdot \bar{h}_1,$$

where g_1, $\bar{g}_1$, h_1, $\bar{h}_1$ are words in the generators $a_{m1}, a_{m2}, \cdots, a_{mm-1}$, and we also know that $g_1 > 1$. We put $t = a_{mm}$. Then

4.33 $$t^{-1} g t \cdot h^{-1} = a_{mm}^{-2} g_1 a_{mm}^{2} \cdot a_{mm}^{-1} \bar{g}_1 h_1^{-1} a_{mm} \cdot \bar{h}_1^{-1} > 1,$$

because $g_1 > 1$; and the result follows.

From this we see immediately:

4.4 Lemma. *If the self-conjugate subgroup $H \subset G_\omega$ contains with any element h also all the elements between h and its inverse,*[20] *then $H = \{1\}$ or $H = G_\omega$.*

This lemma allows us to show that G_ω (in the given order) is what one would call "ordinally simple":

4.5 Theorem. *If G_ω is mapped homomorphically on an ordered group G^* such that $g < h$ in G_ω implies $g^* \leqq h^*$ for the homomorphic images of g and h in G^*, then either the homomorphism is trivial, i. e., $G^* = \{1\}$, or the homomorphism is an isomorphism.*

Proof. Let the kernel of the homomorphism be the self-conjugate subgroup H of G_ω. If $H \neq \{1\}$, $1 < h \in H$, and $1 < g < h$, $g \in G_\omega$ arbitrary, then $1 \leq g^* \leq h^* = 1$; hence $g \in H$. Then $H = G_\omega$ by 4.4, and the homomorphism is trivial. On the other hand, if $H = \{1\}$, then the homomorphism is an isomorphism; which proves the theorem.

University College,
Hull, England.

BIBLIOGRAPHY.

[1] Baer, R., "The higher commutator subgroups of a group," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 143-160.

[2] Birkhoff, G., "Lattice-ordered groups," *Annals of Mathematics* (2) vol. 43 (1942), pp. 298-331.

---

[20] "Symmetric section" (Levi [9]) or "isolated subgroup."

[3] Birkhoff, G., Review of Everett and Ulam, "On ordered groups," *Mathematical Reviews*, vol. 7 (1946).

[4] Everett, C. J., and S. Ulam, "On ordered groups," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 208-216.

[5] Fraenkel, A., *Einleitung in die Mengenlehre*, 3rd ed., Springer, Berlin, 1928.

[6] Hahn, H., "Über die nicht-archimedischen Grössensysteme," *Sitzungberichte der mathematisch-naturwissenschaftlichen Klasse der Akademie der Wissenschaften zu Wien* (IIa), vol. 116 (1907), pp. 601-653.

[7] Hall, P., "A contribution to the theory of groups of prime power order," *Proceedings of the London Mathematical Society* (2), vol. 36 (1933), pp. 29-95.

[8] Hilbert, D., *Grundlagen der Geometrie*, 7th ed., Teubner, Leipzig, 1930.

[9] Levi, F. W., "Ordered groups," *Proceedings of the Indian Academy of Sciences*, vol. 16, (1942), pp. 256-263.

[10] Levi, F. W., "Contributions to the theory of ordered groups," *Proceedings of the Indian Academy of Sciences*, vol. 17 (1943), pp. 199-201.

[11] Steinitz, E., *Algebraische Theorie der Körper* (ed. Baer and Hasse), de Gruyter, Berlin, 1930.

ON UNIFORM SPACES WITH A UNIQUE STRUCTURE.*

By Raouf Doss.

Let E be a separated uniformisable space (see [1] and [2]). We first prove, as a lemma, that if $\mathfrak{F}$ is a filter on E without contiguous point (*point adhérent*), then there exists a uniform structure, compatible with the topology of E, for which $\mathfrak{F}$ is a Cauchy filter.

We propose next to characterize the uniformisable spaces for which there is only one structure. In the space E we shall say that two closed disjoint sets A and B are *normally separable* if there exists a real function $f(x)$, continuous on E, taking the value 0 on A and the value 1 on B. With this definition, in order that the separated uniformisable space E have a unique structure it is necessary and sufficient that of any two normally separable sets one at least be compact.

Lemma. *Let E be a separated uniformisable space and $\mathfrak{F}$ a filter on E without contiguous point. Then there exists a uniform structure $\mathfrak{U}$, compatible with the topology of E, for which $\mathfrak{F}$ is a Cauchy filter.*

Proof. Let a be any element of E and V'_a an arbitrary neighborhood of a. Since a is not contiguous to $\mathfrak{F}$ there exists an $F \in \mathfrak{F}$ such that a is exterior to F. Let W_a be a neighborhood of a contained in V'_a and having no point in common with F and let $f(x)$ be the continuous function taking the value 0 at a and the value 1 outside W_a; in particular $f(x) = 1$ for $x \in F$.

Let $\epsilon > 0$ be arbitrary. Since $f(x)$ is continuous, to every x corresponds a neighborhood V_x^ϵ such that $y \in V_x^\epsilon$ implies $|f(x) - f(y)| < \epsilon$. We shall take V_x^ϵ to be the set of points y for which $|f(x) - f(y)| < \epsilon$. Put

$$V_\epsilon = \bigcup_{x \in E} V_x^\epsilon \times V_x^\epsilon.$$

We have $F \times F \subset V_\epsilon$. In fact, let $b \in F$. Then $f(b) = 1$ and $f(y) = 1$ for $y \in F$. According to the definition of V_b^ϵ we have $F \subset V_b^\epsilon$ i. e., $F \times F \subset V_\epsilon$.

The family of the V_ϵ, $(\epsilon > 0)$, generates a family of *entourages* for the set E. In fact, the V_ϵ are symmetric and the intersection of two V_ϵ is a V_ϵ and contains the diagonal Δ. Moreover, $V_{\epsilon/4} \bigcirc V_{\epsilon/4} \subset V_\epsilon$; for, the relation $(x, y) \in V_{\epsilon/4}$ is equivalent to the existence of a z such that $x \in V_z^{\epsilon/4}$ and $y \in V_z^{\epsilon/4}$; in that case $|f(x) - f(y)| < \epsilon/2$; similarly $(y, z) \in V_{\epsilon/4}$ implies $|f(y) - f(z)| < \epsilon/2$; whence $|f(x) - f(z)| < \epsilon$ and $(x, z) \in V_\epsilon$.

---

* Received January 15, 1948.

Finally, for $\epsilon < 1/2$, $V_\epsilon(a) \subset V'_a$ and $V_\epsilon(x)$ is a neighborhood of x. In fact, $V_\epsilon(a)$ is the set of all x for which $(a, x) \varepsilon V_\epsilon$. For such an x, $|f(a) - f(x)| < 2\epsilon < 1$; whence $f(x) < 1$, $x \varepsilon W_a$ and $V_\epsilon(a) \subset W_a \subset V'_a$. On the other hand, let $y \varepsilon V_x^\epsilon$; since $x \varepsilon V_x^\epsilon$ we have $(x, y) \varepsilon V_\epsilon$; whence $y \varepsilon V_\epsilon(x)$ and $V_x^\epsilon \subset V_\epsilon(x)$, which shows that the $V_\epsilon(x)$ are neighborhoods of x.

We have started from a fixed point a and a fixed neighborhood V'_a of a (and a corresponding function $f(x)$) to obtain the filter generated by the family of the

$$V_\epsilon = V_\epsilon^{(a, V'_a)}.$$

Consider the set ϕ of these filters obtained from all the points a of E and all the neighborhoods V'_a. Let $\mathcal{U}$ be the filter generated by the union of the filters of ϕ. $\mathcal{U}$ is a filter of *entourages* for the set E. The sets

$$V = V_{\epsilon_1}^{(a_1)} \cap V_{\epsilon_2}^{(a_2)} \cdots \cap V_{\epsilon_k}^{(a_k)} \tag{1}$$

where we write $V_{\epsilon_i}^{(a_i)}$ instead of $V_{\epsilon_i}^{(a_i, V'_{a_i})}$, constitute a base for the filter $\mathcal{U}$.

We shall show that the topology induced by the uniform structure whose filter of *entourages* is $\mathcal{U}$ is exactly the topology of E. In fact, for every x, $V(x)$ is a neighborhood of x; moreover, given the arbitrary neighborhood V'_a of a, there exists a $V_\epsilon^{(a)} \varepsilon \mathcal{U}$ for which $V_\epsilon^{(a)}(a) \subset V'_a$.

We shall show, finally, that $\mathfrak{F}$ is a Cauchy filter for the uniform structure $\mathcal{U}$, i. e. that for every $V \varepsilon \mathcal{U}$ there exists an element of $\mathfrak{F}$ small of order V. It is sufficient to consider a V of the form (1). If, therefore, $F^{(a_1)}, \cdots F^{(a_k)}$ are the elements of $\mathfrak{F}$ for which

$$F^{(a_1)} \times F^{(a_1)} \subset V_{\epsilon_1}^{(a_1)}, \cdots F^{(a_k)} \times F^{(a_k)} \subset V_{\epsilon_k}^{(a_k)},$$

then the set $F = F^{(a_1)} \cap \cdots \cap F^{a_k} \varepsilon \mathfrak{F}$ is small of order V, for it is small of order $V_{\epsilon_1}^{(a_1)}, \cdots V_{\epsilon_k}^{(a_k)}$. The lemma is now proved.

We shall study now the uniformisable spaces which have one structure compatible with their topology.

THEOREM. *In order that the separated uniformisable space E have a unique structure, it is necessary and sufficient that of any two normally separable sets one at least be compact.*

Proof. Necessity. If E has a unique structure $\mathcal{U}$, this must coincide with the uniform structure of Weil (see [1], p. 16) for which every continuous function taking values in a uniform space is uniformly continuous. Suppose that A and B are two closed, non-compact, normally separable sets and let

$f(x)$ be the function mentioned in the definition. Since A is not compact there exists on A a filter $\mathcal{F}$ without contiguous point in A. $\mathcal{F}$ is also without contiguous point in E, for, the closures of the sets of $\mathcal{F}$ in A and in E are identical, since A is closed in E. Similarly, there exists on B a filter $\mathcal{G}$ without contiguous point in E. The filter $\mathcal{H}$ generated by the sets H of the form $H = F \cup G$, with $F \varepsilon \mathcal{F}$ and $G \varepsilon \mathcal{G}$, is also without contiguous point. By the lemma, $\mathcal{H}$ is a Cauchy filter for the unique structure $\mathcal{U}$. $f(x)$ being uniformly continuous, the image of $\mathcal{H}$ by f must be the base of a Cauchy filter in the space of the real numbers. But this is not so, for, the image of $\mathcal{H}$ by f has at least two contiguous points 0 and 1. The existence of the two non-compact sets A and B is thus impossible.

Sufficiency. We shall show first, that if the condition of the theorem is satisfied then E is *precompact* for each of its structures. In fact, if E is not precompact for a structure $\mathcal{U}$ there exists an *entourage* $U_0 \varepsilon \mathcal{U}$ such that there is no finite covering of E by sets small of order U_0. (We shall consider only symmetric *entourages*). This means that we can find an enumerable sequence $x_1, x_2, \cdots x_n, \cdots$ of points of E such that we never have $(x_n, x_m) \varepsilon U_0$ for $m \neq n$. Let U_1 be such that $U_1 \bigcirc U_1 \subset U_0$; we conclude that the sets $U_1(x_m)$, $U_1(x_n)$ are disjoint for $m \neq n$. Take U_2 such that $U_2 \bigcirc U_2 \subset U_1$. The neighborhood $U_2(x_n)$ of x_n contains an open neighborhood G_n of x_n and G_n contains a closed neighborhood F_n of x_n. Let $f_n(x)$ be the real continuous function such that $f_n(x_n) = 1$ and $f_n(x) = 0$ for $x \varepsilon CF_n$ and let

$$f(x) = \sum_{n=1}^{\infty} f_{2n}(x).$$

Then $f(x)$ is continuous. In fact, at a point $x \varepsilon G_{2n}$, $f(x)$ is continuous since it coincides with $f_{2n}(x)$ in some neighborhood of x. (G_{2n} is open and the G_{2n} are disjoint). For any x, in particular for $x \varepsilon C \bigcup_n G_{2n}$, $U_2(x)$ has points in common with one $U_2(x_{2n})$ at most. In fact, if, for a certain x_{2n} and a certain y we have $y \varepsilon U_2(x_{2n})$ and $y \varepsilon U_2(x)$, then $(x_{2n}, x) \varepsilon U_1$. If the same were true for another x_{2m} it would give $(x_{2m}, x) \varepsilon U_1$, whence $(x_{2n}, x_{2m}) \varepsilon U_0$, which is excluded. Thus $U_2(x)$ has points in common with one $U_2(x_{2n})$ at most, say with $U_2(x_{2n_0})$. If $x \varepsilon C \bigcup G_{2n}$, x is exterior to F_{2n_0} and there exists a neighborhood H of x such that $H \subset U_2(x)$ and such that H and F_{2n_0} are disjoint. This neighborhood H meets no F_{2n}. We conclude that in H the function f is always zero. Since H is a neighborhood of x the function f is continuous at x.

P14910

If we put

$$A = \{x_1, x_3, \cdots x_{2n-1}, \cdots\}$$

$$B = \{x_2, x_4, \cdots x_{2n}, \cdots\},$$

then $f(x) = 0$ for $x \varepsilon A$ and $f(x) = 1$ for $x \varepsilon B$. If we show that A and B are closed and non-compact this would contradict the assumption of the theorem and we should have proved that E is precompact for each of its structures. It is almost evident that A, for example, is closed, for if $x \bar{\varepsilon} A$, then, as before, $U_2(x)$ would have points in common with one $U_2(x_{2n-1})$ at most and would contain one element of A at most; whence $x \bar{\varepsilon} \bar{A}$. Similarly, A is not compact, for, the sequence $x_1, x_3, \cdots x_{2n-1}, \cdots$ has no accumulation point in A.

Suppose now that E had two different structures $\mathcal{U}$ and $\mathcal{U}'$. It is impossible that $E = \hat{E}(\mathcal{U})$ or $E = \hat{E}(\mathcal{U}')$, ($\hat{E}(\mathcal{U})$ denotes the separated completed space of E for the structure $\mathcal{U}$ and we identify E with the everywhere dense subspace of $\hat{E}(\mathcal{U})$ with which it is isomorphic). In fact, in such a case, E would be compact and would have one structure only. Hence $\hat{E}(\mathcal{U})$ and $\hat{E}(\mathcal{U}')$ have each at least one point more than E. It is impossible that $\hat{E}(\mathcal{U})$ and $\hat{E}(\mathcal{U}')$ have each one point only (ω and ω' respectively) more than E. In fact, in such a case, $\mathcal{U}$ and $\mathcal{U}'$ would be identical. To see this, let (V_ω) be the trace on E of the filter of neighborhoods of ω in $\hat{E}(\mathcal{U})$; this filter has no contiguous point in E but it must have at least one contiguous point in $\hat{E}(\mathcal{U}')$; the only possible point is ω'; since $\hat{E}(\mathcal{U}')$ is compact, we conclude that (V_ω) converges to ω' i. e. is finer than the trace $(V_{\omega'})$ on E of the filter of neighborhoods of ω' in $\hat{E}(\mathcal{U}')$; in the same manner $(V_{\omega'})$ is finer than (V_ω). The two compact spaces $\hat{E}(\mathcal{U})$ and $\hat{E}(\mathcal{U}')$ are thus homeomorphic, which implies that they are isomorphic; the equality of the two structures $\mathcal{U}$ and $\mathcal{U}'$ follows.

Hence we must suppose that $\hat{E}(\mathcal{U})$, for example, has at least two points a and b more than E. Let V'_a and V'_b be two closed disjoint neighborhoods of a and b and let V_a and V_b, respectively, be their traces on E. $\hat{E}(\mathcal{U})$ being compact, hence normal, there exists a real continuous function taking the value 0 on V'_a and the value 1 on V'_b. Its restriction to E takes the value 0 on V_a and the value 1 on V_b. V_a and V_b which are closed in E are thus normally separable. If we show that neither V_a nor V_b is compact, this would contradict the assumption of the theorem and the existence of the two different structures $\mathcal{U}$ and $\mathcal{U}'$ will be impossible.

If V_a were compact it would be closed in $\hat{E}(\mathcal{U})$. But we have $a \bar{\varepsilon} V_a$ and $a \varepsilon \bar{V}_a$. To see that a is contiguous to V_a in $\hat{E}(\mathcal{U})$ it will be sufficient to consider the neighborhoods V''_a of a contained in V'_a. All these neighborhoods meet E since E is everywhere dense in $\hat{E}(\mathcal{U})$, so that all of them meet $V_a = V'_a \cap E$. Hence V_a cannot be compact and the theorem is proved.

Remarks. 1. We deduce easily from what precedes that if E has a unique structure then E is *locally compact.* In fact we have seen that when E is not compact, its unique structure $\mathcal{U}$ must be such that $\hat{E}(\mathcal{U})$ has one point only, ω, more than E. Let $x \varepsilon E$ and let V_x be a closed neighborhood of x in $\hat{E}(\mathcal{U})$, not containing ω. V_x is compact since $\hat{E}(\mathcal{U})$ is compact; its restriction to E is still V_x, which proves that x has a compact neighborhood.

2. In the case of a *metrizable space* E the existence of a unique structure implies that *E is compact.* In fact, by a theorem of J. Dieudonné, [3], E has a uniform structure $\mathcal{U}$ for which it is complete: $E = \hat{E}(\mathcal{U})$. Since $\hat{E}(\mathcal{U})$ must be compact, then E is compact.

FAROUK I UNIVERSITY,
ALEXANDRIA, EGYPT.

BIBLIOGRAPHY.

[1] A. Weil, "Sur les espaces à structure uniforme et sur la topologie générale," *Actualités Scientifiques*, No. 551, Hermann, Paris (1937).

[2] N. Bourbaki, "Topologie générale," Chapters I and II, *Actualités Scientifiques*, No. 858, Hermann, Paris (1940).

[3] J. Dieudonné, *Comptes Rendus de l'Académie des Sciences, Paris,* vol. 209 (1939), p. 108.

ON THE ACCESSORY PARAMETERS OF A FUCHSIAN DIFFERENTIAL EQUATION.*

By Zeev Nehari.

It was shown by Schwarz in his classical memoir on the hypergeometric series [1] [1] that the ratio $w = y_1/y_2$ of two linearly independent solutions y_1 and y_2 of the Fuchsian equation

$$(1) \qquad y'' + \left\{ \sum_{\mu=1}^{n-3} \frac{1 - \alpha_\nu}{z - a_\nu} \right\} y' + \frac{\sum\limits_{\mu=1}^{n-3} \lambda_\mu z^{\mu-1} + A z^{n-3}}{\prod\limits_{n-1}^{\nu=1} (z - a_\nu)} y = 0$$

(all constants real) effects the conformal mapping of the half-plane $\mathfrak{J}\{z\} > 0$ onto a curvilinear polygon, composed of n circular arcs, in the w-plane. $\pi\alpha_1, \cdots, \pi\alpha_{n-1}, \pi\alpha_n$ are the angles of the polygon, where α_n is defined by

$$\alpha_n = \alpha'_n - \alpha''_n, \qquad A = \alpha'_n \alpha''_n, \qquad \alpha'_n + \alpha''_n + \sum_{\nu=1}^{n=1} \alpha_\nu = n - 2,$$

and the points $z = \alpha_\nu$ are mapped onto the vertices with angles α_ν. The $n - 3$ real parameters $\lambda_1, \cdots, \lambda_{n-3}$—called by Klein the "accessory parameters" of equation (1)—correspond to $n - 3$ geometrical variables determining a polygon with given angles in a unique manner (apart from a linear transformation of the w-plane).

The object of this paper is to investigate the functional relationship between the geometrical variables defining the curvilinear polygon and the accessory parameters $\lambda_1, \cdots, \lambda_{n-3}$ of equation (1). A number of particular cases of this problem were discussed—either with the help of Hilbert's theory of integral equations or using Klein's method of continuity—by Hilbert [2], Koenig [3], Klein [4], Hilb [5] and others. In all these cases the main geometrical condition imposed on the curvilinear polygon is the existence of an orthogonal circle, a condition fundamental in the theory of automorphic functions. Another case ($n = 4$, special values of the angles and an additional geometrical condition) was recently treated by v. Koppenfels [6].

We shall begin with the discussion of the case $n = 4$ in which the

---

* Received, July 1, 1947; revised January 21, 1948.

[1] The numbers in brackets refer to the bibliography at the end of the paper.

required analysis is particularly simple but which nevertheless shows all the characteristic features of the general case.

1. The case $n = 4$. Equation (1) reduces in this case to

$$(2)\quad y'' + \left\{ \frac{1-\alpha}{z-a} + \frac{1-\beta}{z-b} + \frac{1-\gamma}{z-c} \right\} y' + \frac{Az+B}{(z-a)(z-b)(z-c)} y = 0$$

with

$$\alpha + \beta + \gamma + \delta' + \delta'' = 2, \qquad A = \delta'\delta'',$$

and obvious changes of notation; B is the accessory parameter. (2) is sometimes called "Heun's equation"; it has recently been studied—from a different point of view—by Svartholm [7] and Erdélyi [8].

The angles of the quadrilateral upon which $\mathfrak{J}\{z\} > 0$ is mapped by the ratio $w(z) = y_1/y_2$ of two linearly independent solutions y_1 and y_2 of (2) are $\pi\alpha$, $\pi\beta$, $\pi\gamma$, $\pi\delta$ respectively, where $\delta = \delta' - \delta''$; the vertices are situated at the points $w(a)$, $w(b)$, $w(c)$, $w(\infty)$. In order to keep the discussion as simple as possible, we shall assume that $0 < \alpha, \beta, \gamma, \delta < 1$, thus confining ourselves to polygons with corners pointing outwards. Unless a restriction of this kind is introduced, there exists quite a surprising variety of different types of curvilinear quadrilaterals, as was shown in detail by Schoenfliess [9] and Ihlenburg [10].

Before we proceed to formulate a statement describing the general functional dependence of the accessory parameter B on the geometry of the quadrilateral, we shall introduce a few notations needed in the sequel. The full circle, part of which coincides with the conformal image of the stretch $a < z < b$, will be called K_{ab}, and similarly for the other circles. We shall characterize the relative position of K_{ab} and $K_{c\infty}$ by the cross-ratio of the four points in which the two circles are cut by the straight line connecting their centers. As there are six possible values of this cross-ratio, we shall have to choose one particular value in order to make things definite. This will be done in the following way:

Among these six values there are two, say θ and θ^*, which become infinite if one of the two circles K_{ab}, $K_{c\infty}$ reduces to a point. These values obviously satisfy $\theta + \theta^* = 1$. It is further easy to see that in case K_{ab} and $K_{c\infty}$ do not intersect, one of the cross-ratios is negative and the other > 1; if K_{ab} and $K_{c\infty}$ intersect, both cross-ratios are positive, one of them—say θ—satisfying $0 < \theta \leq \frac{1}{2}$ and the other $\frac{1}{2} \leq \theta^* < 1$. If K_{ab} and $K_{c\infty}$ touch, the cross-ratios reduce to 0 and 1 respectively and if K_{ab} and $K_{c\infty}$ are orthogonal, both θ and

θ^* reduce to $\frac{1}{2}$. In order to define the cross-ratio in question in a unique manner, we shall first assume that K_{ab} and $K_{c\infty}$ do not intersect. Then there are two possibilities: Either we can, by moving continuously the four vertices on the circles K_{ab} and $K_{c\infty}$ respectively and varying continuously the four angles, convert the given quadrilateral into another quadrilateral (likewise composed of four circular arcs) with four right angles, or we cannot. If this continuous reduction to a "rectangle" is possible, we shall take the negative cross-ratio; in the other case we shall take the positive. If K_{ab} and $K_{c\infty}$ intersect, there are two ways of detaching them from each other in a continuous manner: Either the two circles will pass through a position in which they are orthogonal to each other, or they will not. If we take the second alternative and, while detaching the circles from each other, allow any continuous change of the quadrilaterals which leaves its vertices on the respective circles, there are again two possibilities: Either the new quadrilateral thus obtained is continuously reducible to a "rectangle" in the sense indicated above, or else it is not. In the first case we choose that value of the cross-ratio which is between 0 and $\frac{1}{2}$; in the latter case the value of the cross-ratio has to be taken between $\frac{1}{2}$ and 1.

We have thus defined the value of the cross-ratio in a unique manner in all possible cases (provided the angles are all positive and smaller than π). This particular value will be denoted by θ_1. The cross-ratio defined in the same manner with regard to K_{cb} and $K_{\infty a}$ will be denoted by θ_2.

With the help of these notations, the general result we shall establish may be stated as follows:

The ratio $w(z) = y_1/y_2$ *of two linearly independent solutions of* (2) *maps the half-plane* $\mathfrak{J}\{z\} > 0$ *conformally on a curvilinear quadrilateral with angles* $\pi\alpha$, $\pi\beta$, $\pi\gamma$, $\pi\delta$. *If the cross-ratio* θ_1 *is defined as above and the parameter* λ *is introduced by*

$$\begin{aligned}
&\lambda = 4B + ((a+b+c)/3)\{(2-\alpha-\beta-\gamma)^2-\delta^2\} \\
(3)\qquad &+ 2(1-\alpha-\beta-\gamma)\{(a\alpha+b\beta+c\gamma)-\tfrac{1}{3}(a+b+c)(\alpha+\beta+\gamma)\} \\
&+ a\alpha^2+b\beta^2+c\gamma^2+((a+b+c)/3)(\alpha^2+\beta^2+\gamma^2),
\end{aligned}$$

then λ *is a root of the equation*

$$D(\lambda) = \theta_1, \tag{4}$$

where $D(\lambda)$ *is the absolutely converging determinant* $|A_{nv}|$ *defined by*

$$A_{nv} = \frac{i^{n-v}(n-v)\,q^{n-v/2}}{n^2(1-q^{2n-2v})}$$

$$\{\tfrac{1}{4}-\delta^2-q^{n-v}(\tfrac{1}{4}-\gamma^2)+(-1)^{n-v}[\tfrac{1}{4}-\alpha^2-q^{n-v}(\tfrac{1}{4}-\beta^2)]\}\frac{n\neq v}{n\neq 0}$$

$$A_{nn} = 1-(\omega_1/\pi n)^2\{\lambda+(\eta_1/\omega_1)(1-\alpha^2-\beta^2-\gamma^2-\delta^2)\} \quad (n\neq 0) \tag{5}$$

$$A_{0v} = \frac{\pi^2 v i^v q^{v/2}}{1-q^{2v}}\{\tfrac{1}{4}-\beta^2-q^v(\tfrac{1}{4}-\alpha^2)+(-1)^v[\tfrac{1}{4}-\gamma^2-q^v(\tfrac{1}{4}-\delta^2)]\}$$

$$A_{00} = \lambda\omega_1{}^2+\omega_1\eta_1(1-\alpha^2-\beta^2-\gamma^2-\delta^2).$$

Here, $2\omega_1$ *and* $2\omega_2$ (ω_1 *real,* ω_2 *purely imaginary,* $\mathfrak{I}\{\omega_2/\omega_1\}>0$) *are the periods of Weierstrass'* $\wp$*-function defined by*

$$\wp'^2(u) = 4\{\wp(u)-e_1\}\{\wp(u)-e_2\}\{\wp(u)-e_3\}$$

with

$$e_1 = a-(a+b+c/3); \qquad e_2 = b-(a+b+c/3);$$
$$e_3 = c-(a+b+c/3);$$

q *denotes, as usual,* $\exp\{\pi i\omega_2/\omega_1\}$ *and* η_1 *is the increment of Weierstrass'* ζ*-function if* $2\omega_1$ *is added to the argument.*

If θ_2 *is given,* λ *is a root of*

$$D'(\lambda) = \theta_2, \tag{4a}$$

where $D'(\lambda)$ *is obtained from* $D(\lambda)$ *by interchanging* ω_1 *and* ω_2.

If neither θ_1 *nor* θ_2*, but another invariant property of the quadrilateral is given—e. g., the existence of an orthogonal circle—this property can always be expressed by an identical relation between* θ_1 *and* θ_2*. If this relation is of the form*

$$F(\theta_1,\theta_2) = 0,$$

then λ *is a root of the equation*

$$F\{D(\lambda),D'(\lambda)\} = 0. \tag{6}$$

In all these cases, the equations for λ *have an infinity of roots, although not all of them necessarily real. To all real roots correspond quadrilaterals with the required properties. There is always one particular real root which gives rise to a quadrilateral not overlapping itself provided, of course, such a quadrilateral is geometrically possible.*

Before we prove this statement, we shall transform equation (2) by introducing elliptic functions, as in the usual treatment of Lamé's equation. Writing

$$z = z_1 + (a + b + c/3),$$

$$a - (a + b + c/3) = e_1, \qquad b - (a + b + c/3) = e_2,$$

$$c - (a + b + c/3) = e_3 \qquad (e_1 + e_2 + e_3 = 0),$$

we set

$$z_1 = \wp\,(u),$$

$$(dz_1/du)^2 = 4(z_1 - e_1)(z_1 - e_2)(z_1 - e_3),$$

the elliptic function $\wp\,(u)$ having one real and one purely imaginary period.

We then write

$$y = s(u)v(u)$$

with a suitably chosen $s(u)$ so as to make $v(u)$ satisfy an equation of the form

$$v'' + r(u)v = 0, \tag{7}$$

the primes now denoting differentiations with respect to u. Since $z_1 = \wp\,(u)$ maps a rectangle on the half-plane $\mathfrak{I}\{z_1\} > 0$ and

$$y_1/y_2 = v_1/v_2,$$

where v_1 and v_2 are two linearly independent solutions of (7), the function

$$w = f(u) = v_1(u)/v_2(u)$$

will map a rectangle on a curvilinear quadrangle, the corners of the rectangle corresponding to the vertices of the quadrangle.

By carrying out the required calculations it is found that the exact form of (7) is

$$\begin{aligned} v'' + \{(\tfrac{1}{4} - \delta^2)\wp\,(u) + (\tfrac{1}{4} - \alpha^2)\wp\,(u - \omega_1) + (\tfrac{1}{4} - \gamma^2)\wp\,(u - \omega_2) \\ + (\tfrac{1}{4} - \beta^2)\wp\,(u - \omega_3) + \lambda\}v = 0, \end{aligned} \tag{8}$$

where λ, the only constant remaining undetermined, is the accessory parameter.

This parameter is, of course, not identical with the accessory parameter B of equation (2). By equating the first two coefficients in the expansions of the solutions of (2) and (8) in the vicinity of one of the singular points and remembering the transformation connecting z and u, λ and B are found to be connected by (3).

In order to remove the singularities from the real and imaginary axes—

for reasons which will become clear presently—we make one last transformation, writing $u = t + (\omega_1 + \omega_2/2)$ and $v(u) = \eta(t)$. (8) thus becomes

$$\eta'' + \{(\tfrac{1}{4} - \alpha^2)\wp(t - \omega_1/2 + \omega_2/2) + (\tfrac{1}{4} - \beta^2)\wp(t - \omega_1/2 - \omega_2/2) + (\tfrac{1}{4} - \gamma^2)\wp(t + \omega_1/2 - \omega_2/2) + (\tfrac{1}{4} - \delta^2)\wp(t + \omega_1/2 + \omega_2/2) + \lambda\}\eta = 0, \tag{9}$$

which, for short, we shall write

$$\eta'' + \{g(t) + \lambda\}\eta = 0. \tag{9a}$$

We shall now assume that the circles K_{ab} and $K_{c\infty}$ do not intersect (the case in which they do will be dealt with later). In this case, we can find a linear substitution transforming K_{ab} and $K_{c\infty}$ into two concentric circles with their common center at the origin. The radii of these two circles will be denoted by ρ and r $(\rho < r)$.

We then choose the two independent solutions η_1 and η_2 of (9) in such a way as to make the ratio

$$w(t) = \eta_1/\eta_2 \tag{10}$$

map the fundamental rectangle on the particular quadrilateral thus obtained.

Two sucessive inversions of the rectangle along the two sides parallel to the imaginary axis will result in a point t being shifted by $2\omega_1$; two corresponding inversions with respect to the two concentric circles of radii ρ and r respectively will transform w into $(\rho/r)^2 w$ (or $(r/\rho)^2 w$, if the order of inversions is interchanged). By Schwarz's symmetry principle, we have therefore

$$w(t + 2\omega_1) = (\rho/r)^2 w(t). \tag{11}$$

On the other hand, $g(t)$ having the period $2\omega_1$, equation (9a) is unaffected by replacing by $t + 2\omega_1$. Both $\eta_1(t + 2\omega_1)$ and $\eta_2(t + 2\omega_1)$ will therefore also be solutions of (9a) and consequently be of the form

$$\eta_1(t + 2\omega_1) = A\eta_1(t) + B\eta_2(t)$$
$$\eta_2(t + 2\omega_1) = C\eta_1(t) + D\eta_2(t)$$

with constant A, B, C, D. In view of (10) and (11) this yields

$$\frac{A\eta_1(t) + B\eta_2(t)}{C\eta_1(t) + D\eta_2(t)} = (\rho/r)^2(\eta_1(t)/\eta_2(t)).$$

This is only possible if we have

CALCUTTA CENTRAL LIBRARY.

$$B = C = 0$$
$$A/D = (\rho/r)^2.$$

Now any two solutions of (9a) are connected by a relation

$$\eta_1(t)\eta'_2(t) - \eta_2(t)\eta'_1(t) = \text{const.}$$

Substituting $t + 2\omega_1$, for t, this leads to

$$AD - BC = 1,$$

whence, in view of $B = C = 0$, $A/D = (\rho/r)^2$,

$$A = \pm \rho/r,$$
$$D = \pm r/\rho.$$

It is clear that although both the positive and negative values of A and D so far fulfill all the conditions, one set must be discarded in each particular case, as otherwise we would be led to four linearly independent solutions of a linear differential equation of the second order, which is absurd. In order not to interrupt the argument, we shall postpone the decision as to the correct sign to take, and shall, for the time being, carry both signs in the formulas.

The two solutions singled out by our particular location of the quadrilateral have thus the following transformation properties:

$$\begin{aligned} \eta_1(t + 2\omega_1) &= \pm (\rho/r)\eta_1(t) \\ \eta_2(t + 2\omega_1) &= \pm (r/\rho)\eta_2(t), \end{aligned} \tag{11a}$$

i. e. they are the well-known multiplicative solutions forming the subject of Floquet's theory of differential equations with periodic coefficients. Their further discussion will accordingly be modelled on Hill's treatment of his equation with periodic coefficients occurring in the lunar theory.

If s is one of the solutions of

$$e^s = \pm \rho/r,$$

the function

$$e^{(s/2\omega)t}$$

is also multiplied by $\pm \rho/r$ if t is replaced by $t + 2\omega_1$; $\eta_1(t)$ is therefore of the form

$$\eta_1(t) = e^{(s/2\omega_1)t}\phi(t), \tag{12}$$

where $\phi(t)$ is a periodic function with the period $2\omega_1$, i. e.

$$\phi(t + 2\omega_1) = \phi(t).$$

The function $g(t)$ in (9a) is regular in the strip $-\Im\{\omega_2/2\} < \Im\{t\} < \Im\{\omega_2/2\}$ and the same is therefore true of any solution of (9a). Hence, both $g(t)$ and $\phi(t)$—having the period $2\omega_1$—may be expanded into Fourier series converging absolutely in the interior of this strip:

$$g(t) = \sum_{n=-\infty}^{\infty} c_n e^{(\pi i n/\omega_1)t},$$

$$\phi(t) = \sum_{n=-\infty}^{\infty} a_n e^{(\pi i n/\omega_1)t}.$$

By (12), we have further

$$\eta_1(t) = \sum_{n=-\infty}^{\infty} a_n e^{(\pi i n/\omega_1 + s/2\omega_1)t}.$$

Inserting these expressions in (9a) and equating the coefficients of the powers of $\exp\{(\pi i/\omega_1)t\}$ to zero, we obtain the system of equations

$$a_n\{(\pi i n/\omega_1 + s/2\omega_1)^2 + \lambda + c_0\} + \sum_{\nu=-\infty}^{\infty}{}' a_\nu c_{n-\nu} = 0, \quad n = \cdots, -1, 0, 1, \cdots,$$

where the prime indicates that the term involving a_n has to be omitted from the summation. Writing for short

$$is/2\pi = \sigma, \qquad (\omega_1/\pi)^2(\lambda + c_0) = \mu, \qquad (\omega_1/\pi)^2 c_\nu = d_\nu,$$

this becomes

$$a_n\{(n - \sigma)^2 - \mu\} - \sum_{\nu=-\infty}^{\infty}{}' a_\nu d_{n-\nu} = 0.$$

In order that this system of homogeneous linear equations be consistent, its determinant must vanish provided, of course, it converges. Convergence being easily secured by dividing the n-th equation by $(n - \sigma)^2 - \mu$, σ and μ are thus connected by the equation

$$D(\sigma, \mu) \equiv | A_{n\nu} | = 0,$$

where the elements of the determinant $| A_{n\nu} |$ are defined by

$$A_{n\nu} = -d_{n-\nu}/(n - \sigma)^2 - \mu, \qquad n \neq \nu,$$

$$A_{nn} = 1.$$

The functional dependence of this determinant on the parameter σ can easily found by an artifice going back to Hill [11]. Using his procedure, it is found that the equation $D(\sigma, \mu) = 0$ is equivalent to the equation

$$D_1(\mu) = \sin^2 \pi\sigma,$$

where $D_1(\mu)$ denotes the determinant $B_{n\nu}$ defined by

$$B_{n\nu} = -d_{n-\nu}/n^2 \qquad (n \neq \nu, n \neq 0)$$

$$B_{nn} = 1 - (\mu/n^2) \qquad (n \neq 0)$$

$$B_{0\nu} = \pi^2 d_{-\nu}, \qquad B_{00} = \pi^2 \mu.$$

σ having been defined by

$$\sigma = is/2\pi, \qquad e^s = \pm \rho/r,$$

we have

$$D_1(\mu) = -(r-\rho)^2/4\rho r,$$

if $e^s = s/r$, or

$$D_1(\mu) = (r+\rho)^2/4\rho r = 1 - [-(r-\rho)^2/4\rho r],$$

if $e^s = -s/r$.

The expressions $-(r-\rho)^2/4\rho r$ and $(r+\rho)^2/4\rho r$ have a simple geometrical interpretation. They are two values of the cross-ratio of the four points at which a straight line passing through the common center of the two circles—of radii ρ and r respectively—intersects these circles. It is further obvious that both these cross-ratios become infinite if either ρ or r reduce to zero. As these cross-ratios are invariant with respect to an arbitrary linear transformation, we can make our result independent of the particular location of the quadrilateral assumed in the course of the proof. Since—in this particular case— $-(r-\rho)^2/4\rho r$ or $(r+\rho)^2/4\rho r$ each coincide with one of the cross-ratios θ and θ^* defined above, μ will therefore always be a solution of

$$D_1(\mu) = \theta$$

or

$$D_1(\mu) = \theta^* = 1 - \theta.$$

In order to decide which of these two equations is to be taken in a given case, we recall that the negative cross-ratio corresponds to the periodicity factor ρ/r while the positive one belongs to the factor $-\rho/r$. We now assume that the fundamental quadrilateral is continuously reducible to a "rectangle" in the manner described further above. This reduction will result in α, β, γ, δ all taking the value $\frac{1}{2}$ and λ tending continuously to a value λ_0. (9) will then take the form

$$\eta'' + \lambda_0 \eta = 0.$$

The two multiplicative solutions of this equation are

$$\eta_1 = e^{\sqrt{-\lambda_0 t}}, \qquad \eta_2 = e^{-\sqrt{-\lambda_0 t}}.$$

In view of

$$\eta_1(t + 2\omega_1) = e^{2\omega_1\sqrt{-\lambda_0}}\eta(t),$$

λ_0 must be negative, since the periodicity factor exp $\{2\omega_1\sqrt{-\lambda_0}\}$ is necessarily real (the two circles are assumed not to intersect). Hence both η_1 and η_2 are multiplied by positive factors if $2\omega_1$ is added to the argument. Now ρ and r were kept constant during our process of continuous reduction; moreover, the periodicity factors being continuous functions of all parameters entering the differential equation, a discontinuous jump from positive to negative values is impossible. Hence, the periodicity factors were positive from the beginning. We have thus proved that in case the quadrilateral is continuously reducible to a rectangle, the negative value of the cross-ratio has to be taken. On the other hand, it is easy to see that in case the periodicity factor is positive (and, consequently, the cross-ratio negative), such a reduction must always be possible. If it is impossible, we shall therefore have to take the positive cross-ratio. Recalling our definition of the cross ratio θ_1, μ will accordingly in all cases be a solution of

$$D_1(\mu) = \theta_1.$$

Using the Fourier expansions of the $\wp$ functions constituting $g(t)$—in (9a)—*viz.*,

$$\wp\,(t + (\omega_1 + \omega_2/2)) = \eta_1/\omega_1 - (\pi/\omega_1)^2 \sum_{n=-\infty}^{\infty}{}' (ni^n q^{n/2}/1 - q^{2n})\, e^{(\pi i n/\omega_1)\, t}$$

and the analogous expressions for $p(t \pm \omega_1/2 \pm \omega_2/2)$, and remembering that $\mu = d_0 + (\omega_1/\pi)^2\lambda$, we have thus established formulas (4) and (5). (4a) follows in exactly the same way, the only difference being that parts of ω_1 and ω_2 have now to be interchanged and the solutions of (9) have to be considered in a strip parallel to the imaginary axis.

So far, we have assumed that the pair of circles forming two opposite sides of the quadrilateral, e. g. K_{ab} and $K_{c\infty}$, do not intersect. If they do, under an angle ϵ, say, the argument has to be but slightly modified. The linear transformation now to be applied to the quadrilateral will be such as to transform the two point of intersection of K_{ab} and $K_{c\infty}$ into the points 0 and ∞ respectively. K_{ab} and $K_{c\infty}$ will then become two straight lines intersecting at the origin under the angle ϵ. Two successive inversions with respect to these two lines will result in a point w being tranformed into a point $we^{\pm 2i\epsilon}$, where the positive sign may without loss of generality be taken. By Schwarz's symmetry principle, we shall therefore have, similarly to (11),

$$w(t+2\omega_1)=e^{2i\epsilon}w(t)$$

and consequently

$$\eta_1(t+2\omega_1)=\pm e^{i\epsilon}\eta_1(t),$$

$$\eta_2(t+2\omega_1)=\pm e^{-i\epsilon}\eta_2(t),$$

which again are the multiplicative solutions of Floquet's theory. The rest of the argument can be repeated all over, the only difference being that instead of the cross-ratios θ and θ^* there now appears $\sin^2\epsilon/2$ or $\cos^2\epsilon/2$ according as the periodicity factors are taken as $+e^{\pm i\epsilon}$ or $-e^{\pm i\epsilon}$.

Now $\sin^2\epsilon/2$ and $\cos^2\epsilon/2$ each coincide with one of the cross-ratios θ and θ^* defined above. If the two circles are orthogonal, both cross-ratios reduce to $\frac{1}{2}$. If the two circles touch, the cross-ratios reduce to 0 and 1 respectively. For reasons of continuity it is therefore clear if the correct value of the cross-ratio is between 0 and $\frac{1}{2}$ and the two circles are detached from each other in a continuous manner without passing a position in which they are orthogonal, the cross-ratio must become negative; if the cross-ratio is between $\frac{1}{2}$ and 1 and the circles are detached in such a manner, the cross-ratio will necessarily become >1. It is therefore seen that the number θ_1 defined above will give in all cases the correct value of the cross-ratio.

As shown in the foregoing, we are able to deduce a transcendental equation for the accessory parameter if one of the geometrical quantities θ_1 or θ_2 is given. This, however, will not always be the case. For instance, in the theory of automorphic functions, the fundamental condition imposed on the quadrilateral is the existence of a circle orthogonal to all its sides. In the most general case, any condition invariant with respect to an arbitrary linear transformation may be given.

Now a particular type of curvilinear quadrilateral with given angles and known "modulus" (i. e. the cross-ratio of the points a, b, c, ∞) is characterized by one and only one invariant condition (there is only one accessory parameter). Accordingly, any invariant condition imposed upon it must be expressible as an identical relation between θ_1 and θ_2; independence of θ_1 and θ_2 would mean the possibility of imposing two independent invariant conditions. If this identical relation is of the form

$$F(\theta_1,\theta_2)=0,$$

the accessory parameter λ will be a solution of the equation

$$F\{D(\lambda),D'(\lambda)\}=0.$$

The exact form of F has to be determined separately in each particular case by elementary geometry.

As an example, we shall consider the case of a quadrilateral with zero angles and an orthogonal circle (this leads to an automorphic function with four "holes"—the most immediate generalization of the elliptic modular function). We may assume that the orthogonal circle coincides with the real axis; this can always be brought about by a suitable linear transformation. Since all angles are equal to zero, all the vertices of the quadrilateral must lie on the orthongonal circle, i. e. on the real axis, and the same will be true of the centers of K_{ab}, K_{bc}, $K_{c\infty}$, $K_{\infty a}$. Both cross ratios θ_1 and θ_2 will therefore refer to the same four points, viz. the four vertices. θ_1 and θ_2 are, however, not identical. By transforming each pair of circles into concentric ones, it is seen that we must have $\theta_1\theta_2 = 1$. λ will therefore be a solution of

$$D(\lambda)D'(\lambda) = 1.$$

As shown by (6), $D(\lambda)$ takes a particularly simple form in this case ($\alpha = \beta = \gamma = \delta = 0$), all non-vanishing terms of the determinant being real.

We now make some remarks as to the general character of the solutions of equations of the type of (4), (4a), or (6).

A determinant

$$| a_{n\nu} + d_{n\nu} | \qquad (\delta_{n\nu} = 0, n \neq \nu; \delta_{nn} = 1)$$

being bounded by

$$\prod_{n=-\infty}^{\infty} (1 + \sum_{n=-\infty}^{\infty} | a_{n\nu} |),$$

it can easily be shown that $D(\lambda)$ is an integral function of λ of order not exceeding $\frac{1}{2}$. Since an integral function of order < 1 cannot have exceptional values, equations (4) and (4a) will have an infinity of solutions. The same will be true of (6), as the relation $F(\theta_1, \theta_2) = 0$—arising from elementary geometrical constructions—as algebraical and the equation determining λ can also in this case be written in the form $P(\lambda) = 0$ where $P(\lambda)$ is an integral function of order $\leq \frac{1}{2}$.

In the general case not all of these solutions are necessarily real. Indeed, it may be shown by elementary examples that an equation of the type $D(\lambda) = \theta$, may have only one real solution.

To each real solution λ there corresponds a pair of solutions of (9) leading to quadrilaterals with the prescribed angles and satisfying the additional geometrical conditions. These solutions differ by the number of times and the particular manner in which the quadrilateral overlaps itself; a full discussion of the various types of quadrilateral obtainable in this way will be found in the papers of Klein and Koenig quoted above. There is one particular value of λ leading to a quadrilateral in the ordinary sense, i. e., one not overlapping itself, if such a quadrilateral is geometrically possible. If there is an infinity of real solutions λ, say $\cdots \lambda_{-1}, \lambda_0, \lambda_1, \cdots$ $(\lambda_{n+1} \geqq \lambda_n)$, some indication as to their asymptotical behavior is given by the fact that

$$\sum_n |\lambda_n|^{-\frac{1}{2}-\epsilon} \qquad (\epsilon > 0)$$

converges, the values λ_n being the zeros of an integral function of order $\leqq \frac{1}{2}$.

2. The general case. We shall now consider the functional relationship between the $n-3$ accessory parameters $\lambda_1, \cdots, \lambda_{n-3}$ appearing in (1), and a set of $n-3$ geometrical variables completely determining the curvilinear polygon mapped by the ratio $w(z) = y_1/y_2$ of two linearly independent solutions of (1) onto the half-plane $\mathfrak{J}\{z\} > 0$. This set of geometrical variables will consist of $n-3$ cross-ratios of the same type as those employed in the case $n = 4$.

In order to define the cross-ratios in question, we shall assume that the real numbers a_ν appearing in (1) are ordered by their magnitude, i. e.,

$$a_1 < a_2 < \cdots < a_{n-1} < \infty .$$

The stretch $a_\nu < z < a_{\nu+1} (1 \leqq \nu \leqq n-2)$ is mapped by $w(z) = y_1/y_2$ on a circular arc forming part of a circle, say K_ν; the ray $a_{n-1} < z < \infty$ is mapped on an arc belonging to a circle K_{n-1}. θ will then denote the cross-ratio of the four points in which the straight line connecting the centers of K_ν and K_{n-1} cuts these circles; the particular value of the cross-ratio to be chosen is defined in a similar way as in the case $n = 4$, the only difference being that the two sides of the reduced "rectangle" not coinciding with K_ν and K_{n-1} will be obtained by letting the other sides of the polygon coalesce into two circular arcs.

As before, we shall not use equation (1) in its rational form, but transform it by the introduction of elliptic functions. However, unlike the case $n = 4$ in which one single transformation was sufficient, we shall have to apply to (2) $n-3$ different transformations, each featuring a different quadruple of singular points a which is being made to correspond by the elliptic function to the four corners of the rectangle of half-periods.

Taking, e. g., $a_1, a_2, a_{n-1}, \infty$, we write

$$z = z_1 + (a_1 + a_2 + a_{n-1}/3),$$

$$a_1 - (a_1 + a_2 + a_{n-1}/3) = e_1 \qquad a_2 - (a_1 + a_2 + a_{n-1}/3) = e_2,$$

$$a_{n-1} - (a_1 + a_2 + a_{n-1}/3) = e_3 \qquad (e_1 + e_2 + e_3 = 0),$$

and set

$$z = \wp(u)$$

$$(dz/du)^2 = 4(z_1 - e_1)(z_1 - e_2)(z_1 - e_3).$$

We then write

$$y = s(u)v(u),$$

where $s(u)$ is to be chosen in such a way as to make $v(u)$ satisfy an equation of the form

$$v'' + r(u)v = 0. \tag{13}$$

The ratio $w = v_1(u)/v_2(u)$ of two linearly independent solutions of this equation will map the rectangle $[0, \omega_1, \omega_3, \omega_2]$ ($2\omega_1$, $2\omega_2$ being the periods of $\wp(u)$, ω_1 real, ω_2 purely imaginary, $\omega_3 = \omega_1 + \omega_2$) on the curvilinear polygon, the four corners of the rectangle corresponding to the vertices of the polygon of indices 1, 2, $n - 1$, n.

Carrying out the necessary calculations, it is found that the function $r(u)$ appearing in (13) is of the form

$$\begin{aligned} r(u) = {} & (\tfrac{1}{4} - \alpha_1^2)\wp(u - \omega_1) + (\tfrac{1}{4} - \alpha_2^2)\wp(u - \omega_3) \\ & + (\tfrac{1}{4} - \alpha_{n-1}^2)\wp(u - \omega_2) + (\tfrac{1}{4} - \alpha_n^2)\wp(u) \\ & + \sum_{\nu=3}^{n-2} (\tfrac{1}{4} - \alpha_\nu^2)\{\wp(u + b_\nu) + \wp(u - b_\nu)\} \\ & + \sum_{\nu=3}^{n-2} \mu_\nu\{\zeta(u + b_\nu) - \zeta(u - b_\nu)\} + \mu, \end{aligned} \tag{14}$$

where b_ν is a solution of

$$a_\nu - (a_1 + a_2 + a_{n-1}/3) = \wp(b_\nu)$$

and $\zeta(u)$ is Weierstrass' ζ-function. The $n - 3$ unspecified constants appearing in (13), *viz.*, $\mu, \mu_3, \mu_4, \cdots, \mu_{n-2}$, may easily be linearly expressed in terms of the constants $\lambda_1, \cdots, \lambda_{n-3}$ in (2) by comparing the expansions of $y(z)$ and $v(u)$ and observing the formulas connecting z and u.

$r(u)$ is a periodic function of u with period $2\omega_1$ and may therefore be expanded into a Fourier series converging absolutely in the interior of a strip, parallel to the real axis, in which $r(u)$ is regular. This is certainly the case for $0 < \mathfrak{J}\{u\} < \mathfrak{J}\{\omega_2\}$. In order, however, to be able to operate on the real axis, we write

$$u = t + (\omega_1 + \omega_2/2), \qquad v(u) = \eta(t), \qquad r(u) = g(t) + \mu,$$

thus replacing (13) by

$$\eta'' + \{g(t) + \mu\}\eta = 0. \tag{15}$$

The Fourier series

$$g(t) = \sum_{n=-\infty}^{\infty} c_n e^{(\pi i n/\omega_1)t}$$

will then converge absolutely for $-\mathfrak{I}\{\omega_2/2\} < \mathfrak{I}\{t\} < \mathfrak{I}\{\omega_2/2\}$.

(15) can now be treated in exactly the same way as (9a). The equivalent of (4) will then be an equation

$$D(\mu, \mu_3, \cdots, \mu_{n-2}) = \theta, \tag{16}$$

where $D(\mu, \mu_3, \cdots, \mu_{n-2})$ is again an absolutely converging determinant of Hill's type, built on the Fourier coefficients, c_ν of $g(t)$. These coefficients c_ν are linear functions of the parameters $\mu_3, \mu_4, \cdots, \mu_{n-2}$ and can easily be written down explicitly with the help of (14) and the well-known Fourier expansions of the functions $\wp$ and ζ. As observed above, the numbers $\mu_3, \cdots, \mu_{n-2}, \mu$ are linear functions of the accessory parameters $\lambda_1, \cdots, \lambda_{n-3}$ of (2). (16) may therefore be replaced by an equation

$$D_1(\lambda_1, \cdots, \lambda_{n-3}) = \theta_1. \tag{17}$$

(17) gives a relation between the accessory parameters $\lambda_1, \cdots, \lambda_{n-3}$ and the geometrical variable θ_1, all the other quantities entering (17) being expressible in terms of the singularities a_ν and the angles $\pi\alpha_\nu$. As in the case $n = 4$, it is easily seen that $D(\mu_3, \cdots, \mu_{n-2}, \mu)$ and, hence, also $D_1(\lambda_1, \cdots, \lambda_{n-3})$ is an integral function of order $\leq \frac{1}{2}$ in all its arguments. Accordingly, if θ_1 and $n - 4$ of the accessory parameters, say $\lambda_2, \cdots, \lambda_{n-3}$, are known, there is a discrete infinity of values λ_1 satisfying (17).

If the process resulting in (17) is repeated with the quadruple a_ν, $a_{\nu+1}$, a_{n-1}, ∞ $(\nu < n - 1)$ taking the place of a_1, a_2, a_{n-1}, ∞, we arrive at an analogous equation

$$D_\nu(\lambda_1, \cdots, \lambda_{n-3}) = \theta_\nu . \tag{18}$$

Since there are altogether $n - 3$ equations (18), they will suffice to determine the possible sets of accessory parameters when the cross-ratios θ_ν are known; as before, we are of course only interested in real solutions of (18), as only these give rise to curvilinear polygons in the ordinary sense. There may be an infinity of sets of real solutions of (18) but there is one and only one set leading to a polygon not overlapping itself, if such a configuration is geometrically possible.

If instead of the cross-ratios θ_ν another set of $n - 3$ invariant conditions characterizing the curvilinear polygon is given, we proceed as follows: In addition to the quadruples of points a_ν, $a_{\nu+1}$, a_{n-1}, ∞ we also consider the sets ∞, a_1, a_k, a_{k+1} $(2 \leq k \leq n - 2)$ and define the cross-ratios θ'_k related to these sets in the same way the cross-ratios θ_ν were defined with respect to $a_\nu, a_{\nu+1}, a_{n-1}, \infty$. By the same procedure as before, we then arrive at a system of equations

$$D'_k(\lambda_1, \cdots, \lambda_{n-3}) = \theta'_k \qquad (k = 2, \cdots, n-2) \tag{19}$$

similar to (18). Now if a set of $n-3$ independent conditions determining the geometrical configuration of the polygon is given, it is clear that there must exist $n-3$ independent identical relations of the type

$$F_k(\theta_1, \theta_2, \cdots, \theta_{n-3}, \theta'_k) = 0, \tag{20}$$

as there cannot exist more than $n-3$ independent cross-ratios. The particular form of the functions F_k has to be determined in each case by elementary geometry. By replacing, in (20), θ_ν and θ'_k by the expressions (18) and (19), we thus arrive at $n-3$ independent equations for the $n-3$ unknowns.

THE HEBREW UNIVERSITY,
JERUSALEM.

BIBLIOGRAPHY.

[1] H. Schwarz, *Gesammelte Mathematische Abhandlungen* II, Berlin (1890), pp. 211-259.

[2] D. Hilbert, " Grundzuege einer allgemeinen Theorie der linearen Integralgleichungen, V," *Goettinger Nachrichten* (1906), pp. 439-480.

[3] R. Koenig, " Anwendung der Integralgleichungen auf ein Problem der Theorie der automorphen Funktionen," *Mathematische Annalen*, vol. 71 (1912), pp. 206-213.

[4] F. Klein, " Bemerkungen zur Theorie der linearen Differentialgleichungen zweiter Ordnung," *Mathematische Annalen*, vol. 64 (1907), pp. 175-196.

[5] E. Hilb, " Ueber Kleinsche Theoreme in der Theorie der linearen Differentialgleichungen," *Mathematische Annalen*, vol. 66 (1909), pp. 215-257; *Mathematische Annalen*, vol. 68 (1910), pp. 24-74.

[6] W. v. Koppenfels, " Konforme Abbildung ausgezeichneter Kreisbogenvierecke," *S.-B. Math.-Nat. Abh. Bayer. Akad. Wiss.* (1943), pp. 327-343.

[7] N. Svartholm, " Die Loesung der Fuchsschen Differentialgleichung zweiter Ordnung durch hypergeometrische Polynome," *Mathematische Annalen*, vol. 116 (1939), pp. 413-421.

[8] A. Erdelyi, " The Fuchsian equation of second order with four singularities," *Duke Mathematical Journal*, vol. 9 (1942), pp. 48-58.

[8] A. Erdélyi, " Certain expansions of solutions of the Heun equation," *Quarterly Journal of Mathematics*, Oxford Ser., vol. 15 (1944), pp. 62-69.

[9] A. Schoenfliess, " Ueber Kreisbogendreiecke und Kreisbogenvierecke," *Mathematische Annalen* vol. 44 (1894), pp. 105-124.

[10] W. Ihlenburg, " Ueber die gestaltlichen Verhaeltnisse der Kreisbogenvierecke," *Goettinger Nachrichten* (1908), pp. 225-230.

[11] See e. g. Whittaker and Watson, *Modern Analysis*, 4th Ed., p. 415.

ON THE MODE OF INCREASE OF THE PRODUCT OF BASIC SETS OF POLYNOMIALS.*

By M. Nassif.

1. The mode of increase of a basic set of polynomials is determined by its order [1] ω and type γ; thus a set of order ω' and type γ' will be of smaller increase than the above one if $\omega' < \omega$ or if $\omega' = \omega$ and $\gamma' < \gamma$. The aim of this note is to give an upper bound for the increase of the product set of two basic sets of polynomials of given increase. Concerning this point of view, two results are already known. The first concerns the product of two simple sets of given order and gives an upper bound for the order of their product.[2]

It states that if the simple sets $\{p_n(z)\}$ and $\{q_n(z)\}$ be of order ω_1 and ω_2 respectively then the product set $\{u_n(z)\}$ of the sets $\{p_n(z)\}$ and $\{q_n(z)\}$ in the given order,[3] is of order not exceeding $\omega_1 + 2\omega_2$. However, to give a more complete solution of the problem in this case we must take into account the type of each constituent set as well as the type of their product. In fact it is here shown that the upper bound for the increase of the product set of two simple sets, is independent of the types of the constituent sets, provided that these types are finite. This result is demonstrated in the following

Theorem I. *Let $\{p_n(z)\}$ and $\{q_n(z)\}$ be simple sets of polynomials of order ω_1 and ω_2 respectively and let each be of finite type. Then the product set $\{u_n(z)\}$ is of increase not exceeding order $\omega_1 + 2\omega_2$ type zero.*

The second result is given by J. M. Whittaker and is concerned with the product of two sets of which the inner set $\{q_n(z)\}$ is of the form $\{(Az + B)^n\}$. In fact Whittaker [4] has shown that if $\{p_n(z)\}$ is of order ω and type γ then the product set $\{u_n(z)\}$ of $\{p_n(z)\}$ and $\{q_n(z)\}$, as given above, is of order ω type $\gamma \mid A \mid^{-1/\omega}$.

This result suggests the study of the increase of the product of two sets

---

* Received October 22, 1947.

[1] J. M. Whittaker, *Interpolatory Function Theory*, Cambridge (1933), Ch. I, p. 11.

[2] M. Nassif, "On the order of the product of simple sets of polynomials," *Proceedings of the Mathematical and Physical Society of Egypt*, vol III, 2 (1946), pp. 43-47.

[3] We shall always take the set $\{u_n(z)\}$ to be the product set of the sets $\{p_n(z)\}$ and $\{q_n(z)\}$ in the given order, unless it is otherwise stated.

[4] *Loc. cit.* (1), Ch. I, p. 12.

the inner one of which is simple and of zero order. Let z^n be expressed by the basic sets $\{p_n(z)\}$ and $\{q_n(z)\}$ in the forms:

$$z^n = \sum_i \pi_{ni} p_i(z), \tag{1.1}$$

and

$$z^n = \sum_i \lambda_{ni} q_i(z). \tag{1.2}$$

Write

$$\omega_n^{(1)}(R) = \sum_i |\pi_{ni}| A_i(R),$$

where $A_n(R) = \max_{|z|=R} |p_n(z)|$, and denote by $\omega_n^{(2)}(R)$ the corresponding expression for $\{q_n(z)\}$, writing $B_n(R)$ for the maximum value of $|q_n(z)|$ in $|z| \leq R$.

Now if $\{p_n(z)\}$ is simple then according to the result given in the above quoted paper, the order of the set $\{u_n(z)\}$ will not exceed that of $\{p_n(z)\}$. In order to give a precise upper bound for the increase of the product set $\{u_n(z)\}$ in this case we have to take the type of $\{p_n(z)\}$ into account and, for the inner set $\{q_n(z)\}$ we consider the behaviour of $\delta^{(2)}(R)$, given by

$$\delta^{(2)}(R) = \overline{\lim_{n\to\infty}} \{\omega_n^{(2)}(R)\}^{1/n}.$$

In fact we shall confine ourselves to the case where $\delta^{(2)}(R)$ is finite for finite values of R; and since $\{q_n(z)\}$ is simple and consequently $\delta^{(2)}(R)/R$ is monotonic decreasing, as given by Whittaker,[5] then $\delta^{(2)}(R) = O(R)$ as R tends to infinity. We shall further impose the condition

$$0 < c \leq \underline{\lim_{n\to\infty}} |q_{nn}|^{1/n} \leq \overline{\lim_{n\to\infty}} |q_{nn}|^{1/n} < \infty. \tag{1.3}$$

For the set $\{p_n(z)\}$ we shall assume that it is a Cannon set. Under these conditions the following theorem can be considered as a generalisation of Whittaker's result:

THEOREM II. *Let $\{p_n(z)\}$ be a Cannon set of polynomials of order ω type γ and let $\{q_n(z)\}$ be a simple set, satisfying (1.3), for which $\delta^{(2)}(R)$ is finite for finite values of R. The product set $\{u_n(z)\}$ will then be of increase not exceeding order ω type $\gamma(k/c)^{1/\omega}$, where $k = \lim_{R\to\infty} \delta^{(2)}(R)/R$.*

The problem is then considered for a wider class of basic sets of polynomials. It is however shown that no upper bound for the order of the product set is obtainable if the order of growth of the coefficients of the polynomials of the inner set $\{q_n(z)\}$ is unrestricted. In fact we impose the restriction:

---

[5] *Series of Polynomials*, Edited by Fouad I University, Cairo (1942), § 4, p. 13.

$$|q_{ni}| = O(n^{\mu n}), \text{ for all } i\text{'s}, \tag{1.4}$$

where μ is a positive finite number. Then the following theorem shows that, with this restriction, no generalisation can be affected in the class of constituent sets beyond the class for which $D_n = O(n)$, where D_n is the degree of the polynomial of highest degree in (1.1) and (1.2).

THEOREM III. *Let* ω_1 *and* ω_2 *be respectively the order of the basic sets* $\{p_n(z)\}$ *and* $\{q_n(z)\}$ *for which* $\overline{\lim\limits_{n\to\infty}} D_n/n = k_1$ *and* k_2 *respectively, and suppose further that* $\{q_n(z)\}$ *satisfies* (1.4). *Then the product set will be of order not exceeding* $k_2\omega_1 + \omega_2 + \mu k_1 k_2$.

2. Proof of Theorem I. Let $f(z) = \Sigma a_n z^n$ be an integral function of order $1/(\omega_1 + 2\omega_2)$ and type α, where α is *any* positive number. We shall prove that the set $\{u_n(z)\}$ represents $f(z)$ in any finite region of the plane.

Now since the set $\{p_n(z)\}$ is of order ω_1 and of finite type then we can choose a number $k_1 > 1$ and a positive integer n_1 such that

$$\omega_n^{(1)}(R) < k_1^n n^{\omega_1 n} \qquad (n > n_1). \tag{2.1}$$

Similarly two numbers $k_2 > 1$ and $n_2 > n_1$ are chosen in such a way that

$$|\lambda_{ni}|\, B_i(R) < \omega_n^{(2)}(R) < k_2^n n^{\omega_2 n} \qquad (n > n_2;\ \text{all } i). \tag{2.2}$$

Also for a number $\alpha_1 > \alpha$, there exists a positive integer $n_3 > n_2$ such that

$$|a_n| < \{\alpha_1 e/n(\omega_1 + 2\omega_2)\}^{n(\omega_1+2\omega_2)} \qquad (n > n_3). \tag{2.3}$$

As $f(z)$ is of order $< 1/\omega_2$ then according to Whittaker [6] the set $\{q_n(z)\}$ represents $f(z)$ in any finite region of the plane in a series of the form $\Sigma b_n q_n(z)$, where by (2.2) and (2.3), b_n can be easily seen to satisfy

$$|b_n|\, B_n(R) < 2k_2^n/n^{n(\omega_1+\omega_2)}\{\alpha_1 e/(\omega_1 + 2\omega_2)\}^{n(\omega_1+2\omega_2)} \quad (n > n_3). \tag{2.4}$$

Also since we can always take q_{nn} to be unity then (2.2) implies that

$$B_n(R) < k_2^n n^{\omega_2 n} \qquad (n > n_2).$$

Hence if $M_n(R)$ is the maximum value of $|u_n(z)|$ in $|z| \leq R$, then, for $R \geq 1$ and for $n > n_2$,

$$M_n(R) < (n+1) S(R) A_n(R) k_2^n n^{\omega_2 n}, \tag{2.5}$$

---

[6] *Loc. cit.* (1), Ch. I, p. 13.

where $S(R)$ is bounded. Writing $\phi_n(R) = \sum_i |\pi_{ni}| M_i(R)$, then by (2.1), (2.4) and (2.5) we obtain, for $n > n_3$ and for $R \geq 1$,

$$|b_n| \phi_n(R) < 2(n+1)T(R)(k_1 \cdot k_2{}^2/R)^n \{\alpha_1 e/(\omega_1 + 2\omega_2)\}^{n(\omega_1+2\omega_2)},$$

where $T(R)$ is bounded. Making n tend to infinity it follows that

$$\overline{\lim_{n\to\infty}} \{|b_n| \phi_n(R)\}^{1/n} \leq k_1 k_2{}^2/R\{\alpha_1 e/(\omega_1 + 2\omega_2)\}^{\omega_1+2\omega_2} < 1, \tag{2.6}$$

for sufficiently large values of R. Since $\phi_n(R)$ is an increasing function of R, (2.6) holds for all positive values of R. Now as it is easy to deduce from the definition of the product set that $q_n(z) = \sum_i \pi_{ni} u_i(z)$ then in view of (2.6) it can be easily shown that the set $\{u_n(z)\}$ represents $f(z)$ in any finite region of the plane.

We now apply Cannon's theorem[7] to deduce from this statement that the product set $\{u_n(z)\}$ will be of increase not exceeding order $\omega_1 + 2\omega_2$ type $1/\alpha$. Finally since α can be taken as large as we please we conclude that the product set is of increase not exceeding order $\omega_1 + 2\omega_2$ type zero, as required.

The following example shows that the upper bound for the increase of the product set $\{u_n(z)\}$ as given by Theorem I is the best possible.

Example I: Consider the basic sets $\{p_n(z)\}$ and $\{q_n(z)\}$ given by

$$p_n(z) = \begin{cases} a^n(n!) + n!z^{n-1}/(\log n)^n + z^n & (n \text{ odd}), \\ z^n & (n \text{ even}). \end{cases}$$

and

$$q_n(z) = \begin{cases} 1 & (n = 0), \\ (3b)^{3n}(n!)^3 + (n!)^3 z^{n-1}/(\log n)^n + z^n & (n \text{ even} > 0), \\ z^n & (n \text{ odd}), \end{cases}$$

where a and b are any positive finite numbers.

It is easily seen that $\{p_n(z)\}$ is of order 1 type a, and $\{q_n(z)\}$ is of order 3 type b.

Forming the product set $\{u_n(z)\}$ we obtain

$$u_n(z) = \begin{cases} 1 & (n = 0) \\ (3b)^{3n}(n!)^3 + (n!)^3 z^{n-1}/(\log n)^n + z^n & (n \text{ even} > 0) \\ a^n(n!) + \dfrac{(3b)^{3(n-1)} \cdot n!\{(n-1)!\}^3}{(\log n)^n} + \dfrac{n!\{(n-1)!\}^3 z^{n-2}}{(\log n)^n\{\log(n-1)\}^{n-1}} \\ \qquad + \dfrac{n!z^{n-1}}{(\log n)^n} + z^n. & (n \text{ odd}). \end{cases}$$

---

[7] B. Cannon, "On the representation of integral functions by general basic series," *Mathematische Zeitschrift*, Bd. 45 (1939), p. 187.

Let

$$z^n = \sum_i \gamma_{ni} u_i(z),$$

and write

$$\Omega_n(R) = \sum_i |\gamma_{ni}| M_i(R). \tag{2.7}$$

Hence for the above set, it is easily seen that

$$\Omega_n(R) = 2a^n n! + \frac{2(3b)^{3(n-1)} n!\{(n-1)!\}^3}{(\log n)^n} + \frac{2n!\{(n-1)!\}^3 R^{n-2}}{(\log n)^n \{\log(n-1)\}^{n-1}} + \frac{2n! R^{n-1}}{(\log n)^n} + R^n,$$

when n is odd, and

$$\Omega_n(R) = \frac{2a^{n-1}(n!)^3(n-1)!}{(\log n)^n} + \frac{2(3b)^{3(n-2)}(n!)^3\{(n-2)!\}^3(n-1)!}{(\log n)^n\{\log(n-1)\}^{n-1}} + \frac{2(n!)^3(n-1)!\{(n-2)!\}^3 R^{n-3}}{(\log n)^n\{\log(n-1)\}^{n-1}\{\log(n-2)\}^{n-2}} + \frac{2(n!)^3(n-1)! R^{n-2}}{(\log n)^n\{\log(n-1)\}^{n-}} + \frac{2(n!)^3 R^{n-1}}{(\log n)^n} + R^n,$$

when n is even > 0. Hence the order Ω of the product set is given by

$$\Omega = \overline{\lim_{n\to\infty}} \log \Omega_n(R)/n \log n = \lim_{\substack{n\to\infty \\ n \text{ even}}} \frac{1}{n\log n} \log \frac{(3b)^{3(n-2)}(n!)^3\{(n-2)!\}^3(n-1)!}{(\log n)^n\{\log(n-1)\}^{n-1}} = 7,$$

and since

$$\lim_{n\to\infty} 1/n^7 \left[\frac{(3b)^{3(n-2)}(n!)^3(n-1)!\{(n-2)!\}^3}{(\log n)^n\{\log(n-1)\}^{n-1}}\right]^{1/n} = 0,$$

then the type of the product set is zero as given by the theorem.

3. Proof of Theorem II. We take $f(z)$ any integral function of order $1/\omega$ type $\sigma < 1/\gamma(c/k)^{1/\omega}$. Then numbers $\gamma_1 > \gamma$, $k_1 > k$, $c_1 < c$ and $\sigma_1 > \sigma$ can be chosen in such a way that

$$\sigma_1 < 1/\gamma_1(c_1/k_1)^{1/\omega}. \tag{3.1}$$

Corresponding to these numbers, positive integers n_1 and $n_2 > n_1$ and a number R_0 exist so that

$$\omega_n^{(1)}(R) < \{n\gamma_1\omega/e\}^{n\omega} \qquad (n > n_1;\ \text{all } R > 0) \tag{3.2}$$

and

$$|\lambda_{ni}| B_i(R) < (k_1 R)^n \qquad (R > R_0,\ n > n_2;\ \text{all } i) \tag{3.3}$$

As in Theorem I, $\{q_n(z)\}$ represents $f(z)$ in any finite region of the plane in a series of the form $\Sigma \beta_n q_n(z)$, where β_n is such that

$$|\beta_n| B_n(R) < 2(k_1R)^n(\sigma_1 e/\omega n)^{\omega n}, \tag{3.4}$$

for $R > R_0$, and for sufficiently large values of n, for $n > n_0$, say.

Also, since the set $\{q_n(z)\}$ is simple then $\lambda_{nn} = 1/q_{nn}$; thus in view of (1.3) and (3.3) an integer $n_3 > n_2$, corresponding to c_1 and a positive number $b < \infty$, exists so that

$$(c_1R)^n < B_n(R) < (bk_1R)^n \qquad (n > n_3;\ R > R_0). \tag{3.5}$$

Let $R_1 > R > R_0$, then (3.5) yields, for $n > n_3$,

$$\begin{aligned} M_n(R) &< \{\sum_{r=0}^{n_3} + \sum_{r > n_3}\} \,|p_{nr}|\, B_r(R) \\ &< A_n(bk_1R_1) S(R, R_1) \end{aligned} \tag{3.6}$$

where $S(R, R_1)$ is bounded. As usual the relations (3.2), (3.4), (3.5) and (3.6) are combined together to yield, for $n > (n_0, n_3)$,

$$|\beta_n| \phi_n(R) < 2T(R, R_1)(k_1/c_1)^n(\sigma_1\gamma_1)^{\omega n},$$

where $T(R, R_1)$ is bounded. In view of (3.1) we conclude that

$$\overline{\lim_{n\to\infty}} \{ |\beta_n| \phi_n(R)\}^{1/n} < 1,$$

and thus the proof of the theorem is completed as in Theorem I.

An example is again constructed to show that the upper bound given in Theorem II is exact.

Example II. Consider the basic sets $\{p_n(z)\}$ and $\{q_n(z)\}$ given by

$$p_n(z) = \begin{cases} z^n & (n \text{ even}), \\ (4a)^{4n}(n!)^4 + (n!)^4 z^{n-1}/(\log n)^n + z^n & (n \text{ odd}), \end{cases}$$

and

$$q_n(z) = \begin{cases} 1 & (n = 0), \\ 1 + 2^n z^{n-1} + z^n & (n \text{ even} > 0), \\ c^n z^n & (n \text{ odd}), \end{cases}$$

where a and c are positive finite numbers and $c < 1$.

It is easily seen that the set $\{p_n(z)\}$ is of order 4 type a, and, for the set $\{q_n(z)\}$, $\delta^{(2)}(R) = 2R$ for $R \geq 1/2$, and

$$0 < c = \underline{\lim_{n\to\infty}} |q_{nn}|^{1/n} < \overline{\lim_{n\to\infty}} |q_{nn}|^{1/n} = 1.$$

The product set $\{u_n(z)\}$ is given by

$$u_n(z) = \begin{cases} 1 & (n=0), \\ 1 + 2^n z^{n-1} + z^n & (n \text{ even} > 0), \\ (4a)^{4n}(n\,!)^4 + \dfrac{(n\,!)^4}{(\log n)^n}\,(1 + 2^{n-1}z^{n-2} + z^{n-1}) + c^n z^n & (n \text{ odd}). \end{cases}$$

It is easily seen by the aid of (2. 7) that, when n is odd,

$$\Omega_n(R) = \frac{2(4a)^{4n}}{c^n}(n\,!)^4 + \frac{2(n\,!)^4}{c^n(\log n)^n} + \frac{2^n}{c^n}\cdot\frac{(n\,!)^4}{(\log n)^n}R^{n-2} + \frac{2(n\,!)^4 R^{n-1}}{c^n(\log n)^n} + R^n,$$

and when n is even > 0,

$$\Omega_n(R) = \frac{2^{n+1}}{c^{n-1}}(4a)^{4(n-1)}\{(n-1)\,!\}^4 + \frac{2^{n+1}\{(n-1)\,!\}^4}{c^{n-1}\{\log(n-1)\}^{n-1}} + \frac{2^{2n-1}\{(n-1)\,!\}^4R^{n-3}}{c^{n-1}\{\log(n-1)\}^{n-1}} + \frac{2^{n+1}\{(n-1)\,!\}^4R^{n-2}}{c^{n-1}\{\log(n-1)\}^{n-1}} + 2^{n+1}R^{n-1} + R^n.$$

On calculating the order Ω and type Γ of the product set $\{u_n(z)\}$ we can easily see that $\Omega = 4$ and $\Gamma = a(2/c)^{1/4}$, as given by Theorem II.

4. The class of simple sets is the only class which admits an upper bound for the increase of the product of two sets of given increase, without restriction on the coefficients of the polynomials of the sets, apart from that governing the coefficient q_{nn}. In fact if $\{q_n(z)\}$ is not simple and of positive order, then we can find instances in which the order of the product set is infinite, even if $q_{nn} = 1$, as it is seen in the following example:

Example III. Consider the basic sets $\{p_n(z)\}$ and $\{q_n(z)\}$ given by

$$p_n(z) = \begin{cases} 1 + z & (n=0), \\ z^n + n^n z^{2n+1} & (n \text{ even} > 0), \\ z^n. & (n \text{ odd}), \end{cases}$$

and

$$q_n(z) = \begin{cases} 1 & (n=0), \\ \dfrac{1}{(n-1)^{3(n-1)}}\,z^{n-1} + z^n & (n \text{ even} > 0), \\ 1 + z & (n=1), \\ z^n + n^{(n^2+3n)}z^{n+1} & (n \text{ odd} > 1). \end{cases}$$

We see that $\overline{\lim_{n\to\infty}}\, D_n/n = 2$ for the set $\{p_n(z)\}$ and is equal to 1 for the set $\{q_n(z)\}$ although this last set is not simple. We also observe that $\{p_n(z)\}$ is of order 1 and $\{q_n(z)\}$ is of order 3. Forming the product set $\{u_n(z)\}$ we obtain

$$u_n(z) = \begin{cases} 2 + z & (n = 0), \\ \dfrac{z^{n-1}}{(n-1)^{3(n-1)}} + z^n + n^n z^{2n+1} + n^n(2n+1)^{\{(2n+1)^2+3(2n+1)\}} z^{2n+2} & (n \text{ even} > 0), \\ 1 + z & (n = 1), \\ z^n + n^{(n^2+3n)} z^{n+1} & (n \text{ odd} > 1). \end{cases}$$

It is easily seen that, when n is odd,

$$\Omega_n(R) = \frac{1 + n^{-n^2}}{1 - n^{-n^2}} R^n + \frac{2n^{3n}}{1 - n^{-n^2}} R^{n+1} + \frac{2n^{3n}(n+1)^{(n+1)}}{1 - n^{-n^2}} R^{2n+3}$$
$$+ \frac{2n^{3n}(n+1)^{(n+1)}(2n+3)^{\{(2n+3)^2+3(2n+3)\}}}{1 - n^{-n^2}} R^{2n+4}.$$

Thus the order of the product set is

$$\Omega \geq \lim_{n\to\infty, n \text{ odd}} \log \Omega_n(R)/n \log n = \infty.$$

Proof of Theorem III. The proof of this theorem is on the same lines as those of the former theorems; our integral function $f(z) = \Sigma a_n z^n$ is taken to be of order $\rho < 1/(k_2\omega_1 + \omega_2 + \mu + k_1 k_2)$. We choose the numbers $\rho_1 > \rho$, $\rho_2 > \omega_1$, $\rho_3 > \omega_2$, $K_1 > k_1$ and $K_2 > k_2$ in such a way that

(4.1) $$\rho_1 < 1/(K_1\rho_2 + \rho_3 + \mu K_1 K_2).$$

Positive integers $n_1 < n_2 < n_3 < n_4 < n_5$ are chosen to satisfy the following inequalities:

(4.2) $$|a_n| < n^{-n/\rho_1} \qquad (n > n_1),$$

(4.3) $$\omega_n^{(1)}(R) < n^{n\rho_2} \qquad (n > n_2),$$

(4.4) $$|\lambda_{ni}| B_i(R) < n^{n\rho_3} \qquad (n > n_3;\ \text{all } i),$$

(4.5) $$D_n < K_1 n \qquad (n > n_4),$$

where D_n is defined for the set $\{p_n(z)\}$, and,

(4.6) $$D'_n < K_2 n \qquad (n > n_5),$$

where D'_n is defined for the set $\{q_n(z)\}$.

Finally in view of (1.4) we choose the number $\alpha > 0$, so that

(4.7) $$|q_{ni}| < \alpha n^{\mu n} \qquad (\text{all } n \text{ and all } i).$$

Now if $f(z)$ is represented by the set $\{q_n(z)\}$ by a series of the form $\Sigma b_n q_n(z)$, then we can deduce from (4. 2), (4. 4) and (4. 6), that

$$(4.8) \qquad |b_n| B_n(R) < 2/(n/K_2 - 1)^{(n/K_2-1)(1/\rho_1-\rho_3)} \qquad (n > K_2 n_5).$$

Also with the aid of (4. 3), (4. 5), (4. 6) and (4. 7) we obtain, for $n > n_5$ and for $R \geq 1$,

$$(4.9) \qquad \begin{aligned} \phi_n(R) &= \sum_i |\pi_{ni}| M_i(R) \\ &< \alpha(K_1 n + 1)(K_1 K_2 n + 1) K_1^{\mu K_1 n} R^{K_1 K_2 n} n^{n(\mu K_1 + \rho_2)}. \end{aligned}$$

Combining (4. 8) and (4. 9) we deduce that, in view of (4. 1),

$$(4.10) \qquad \overline{\lim_{n\to\infty}} \{ |b_n| \phi_n(R) \}^{1/n} \leq \lim_{n\to\infty} \frac{R^{K_1 K_2} K_2^{1/K_2(1/\rho_1-\rho_3)} K_1^{\mu K_1}}{n^{1/K_2(1/\rho_1-\rho_3-K_2\rho_2-\mu K_1 K_2)}} < 1,$$

since we can always take $B_n(R) \geq 1$, for sufficiently large values of R. Since $\phi_n(R)$ is an increasing function of R then (4. 10) holds for all positive values of R. The theorem is then completed in the usual way.

The following example shows that the above result is a best possible one.

Example IV. Consider the basic set $\{q_n(z)\}$ given by

$$q_n(z) = \begin{cases} 1 & (n = 0), \\ \dfrac{1}{(n-1)^{3(n-1)}} z^{n-1} + z^n & (n \text{ even } > 0), \\ 1 + z & (n = 1), \\ z^n + n^{4n} z^{n+1} & (n \text{ odd } > 1). \end{cases}$$

We see that $|q_{ni}| \leq n^{4n}$, and $\overline{\lim_{n\to\infty}} D_n/n = 1$, so that $\mu = 4$ and $k_2 = 1$. The order of this set is easily seen to be equal to 3. Forming the product set $\{u_n(z)\}$ of the set $\{p_n(z)\}$ given in Example III and the set $\{q_n(z)\}$ given above we obtain

$$u_n(z) = \begin{cases} 2 + z & (n = 0), \\ \dfrac{z^{n-1}}{(n-1)^{3(n-1)}} + z^n + n^n z^{2n+1} + n^n (2n+1)^{4(2n+1)} z^{2n+2} & \\ & (n \text{ even } > 0), \\ 1 + z & (n = 1), \\ z^n + n^{4n} z^{n+1} & (n \text{ odd } > 1). \end{cases}$$

Hence in view of (2. 7), $\Omega_n(R)$ is given by

$$\begin{aligned} \Omega_n(R) = 1/\{(n-1)^{(n-1)} - 1\} [2R^{n-1}/(n-1)^{3(n-1)} + \{(n-1)^{(n-1)} + 1\} R^n \\ + 2n^n R^{2n+1} + 2n^n (2n+1)^{4(2n+1)} R^{2n+2}], \end{aligned}$$

when n is even > 2, and

$$\Omega_n(R) = \frac{1 + n^{-n}}{1 - n^{-n}} R^n + \frac{2n^{3n}}{1 - n^{-n}} R^{n+1} + \frac{2n^{3n}(n+1)^{(n+1)}}{1 - n^{-n}} R^{2n+3}$$
$$+ 2\,\frac{n^{3n}(n+1)^{(n+1)}(2n+3)^{4(2n+3)}}{1 - n^{-n}} R^{2n+4},$$

when n is odd > 1. On calculating the order Ω of the set $\{u_n(z)\}$ we obtain

$$\Omega = 12,$$

a result which confirms with the upper bound given in the above theorem.

FAROUK I UNIVERSITY,
ALEXANDRIA, EGYPT.

GAUSS SUMMABILITY OF TRIGONOMETRIC INTEGRALS.*

By S. Bochner and K. Chandrasekharan.

1. Introduction. In the case of an ordinary Fourier series

$$(1) \qquad f(x) \sim \sum a_n e^{inx}$$

the partial sums

$$(2) \qquad S_n(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi}(\sin\,(n+\tfrac{1}{2})t/\sin\tfrac{1}{2}t)f(x+t)\,dt$$

or the customary approximating sums like the Fejér sum

$$(3) \qquad \sigma_n(x) = \frac{1}{2\pi n}\int_{-\pi}^{\pi}(\sin\tfrac{1}{2}nt/\sin\tfrac{1}{2}t)^2 f(x+t)\,dt$$

are finite integrals, and can be written down immediately. The problem of convergence arises only if we want to consider their limit as $n \to \infty$. This is not the case, however with Fourier integrals for instance. If we introduce the Fourier transform

$$(4) \qquad \phi(\alpha) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\alpha x} f(x)\,dx$$

then for the partial sum

$$(5) \qquad S_R(x) = \int_{-R}^{+R} e^{i\alpha x}\phi(\alpha)\,d\alpha$$

we have

$$(6) \qquad S_R(x) = \int_{-\infty}^{+\infty}(\sin Rt/t)f(x+t)\,dt$$

and for the Fejér sum

$$(7) \qquad \sigma_R(x) = \int_{-R}^{+R}(1-(|\,\alpha\,|)/R))e^{i\alpha x}\phi(\alpha)\,d\alpha$$

we have

$$(8) \qquad \sigma_R(x) = \int_{-\infty}^{+\infty}(\sin Rt/t)^2 f(x+t)\,dt.$$

The expressions (5) and (7) are finite integrals, but the expressions (6) and (8) already involve convergence requirements on $f(t)$. The difficulties thus encountered may, however, be said to be due mainly to the "behavior of $f(x)$ at infinity." In fact, the very process of introducing the transform (4)

---

* Received January 6, 1948.

requires some restrictions on $f(x)$, and such restrictions are usually sufficient to give convergence and validity to the integrals (6) and (8) for every finite R.

However, in somewhat different situations, there is another source of difficulty applying only to expressions of the type (6) or (8) and not at all to the analogues of the integrals (4) themselves, and these difficulties stem from the spectral structure proper of the function rather than only from a variable behavior at infinity. This phenomenon was encountered by Bochner [1, 2] when generalizing the classical theory from periodic to almost periodic functions. If

$$f(x) \sim \sum a_{\lambda_n} e^{i\lambda_n x} \tag{9}$$

is an almost periodic function in the sense of Bohr, in which case the Fourier coefficients are determined by the formula

$$a_{\lambda_n} = \lim_{T\to\infty} 1/2T \int_{-T}^{+T} f(t)\, e^{i\lambda_n t} dt \tag{10}$$

then for any given $R > 0$, there may or may not exist the 'Dirichlet-partial-sum,'

$$S_R(x) \sim \sum_{|\lambda_n| \leq R} a_{\lambda_n} e^{i\lambda_n x}; \tag{11}$$

in other words, there may or may not exist an almost periodic function $S_R(x)$ whose Fourier expansion is the series on the right of (11). However, there always does exist the Fejér sum [1]

$$\sigma_R(x) \sim \sum_{|\lambda_n| \leq R} (1 - (|\lambda_n|)/R))\, a_{\lambda_n} e^{i\lambda_n x}, \tag{12}$$

and it can be represented by the integral (8) which is absolutely convergent for any (bounded) almost periodic function. The integral (6) is obviously a 'delicate' one, since it is not absolutely convergent for a bounded function, and indeed as already stated, the partial sum (11) need not 'exist' in cases in which the spectral point $\lambda = R$ or $\lambda = -R$ (or both) is an accumulation point of the Fourier exponents λ_n occurring in the expansion (9). But even if the function (11) does exist, as it very emphatically does in the case of a pure periodic function (1), its integral representation (6) still remains 'delicate,' and the problem arises of securing its convergence and validity in some appropriate manner. This problem was recently attacked by Bochner in [3], and it was shown that for pure periodic (1) and some other almost periodic functions (9), the formula (6) is valid in the following version

---

[1] This sum must not be confused with the so-called Bochner-Fejér sums for almost periodic functions, which are constructed in an entirely different manner.

$$\sigma_R(x) = \lim_{\epsilon \to 0} \int_{-\infty}^{+\infty} e^{-\epsilon^2 t^2} (\sin Rt/t) f(x+t)\, dt \tag{13}$$

which amounts to a Gauss summability of the original version (6). Actually the result was stated for spherical summability of multiple Fourier series.

If

$$f(x_1, \cdots, x_k) \sim \sum a_{n_1 \ldots n_k} e^{i(n_1 x_1 + \ldots + n_k x_k)}, \tag{14}$$

and

$$S^\delta(x; R) \equiv S^\delta(R) \equiv \sum_{\nu^2 = n_1^2 + \ldots + n_k^2} (1 - (\nu^2/n^2))^\delta a_{n_1 \ldots n_k} e^{i(n_1 x_1 + \ldots + n_k x_k)}, \tag{15}$$

and

$$\begin{gathered} f_0(x; t) \equiv f_0(t) = 1/(2\pi)^{k/2} \int_\sigma f(x_1 + \xi_1 t, \cdots, x_k + \xi_k t)\, dw_\xi, \\ 0 < t < \infty, \end{gathered} \tag{16}$$

where σ denotes the unit sphere $\xi_1^2 + \cdots + \xi_k^2 = 1$, and dw_ξ is the $k-1$ dimensional volume-element (if $k = 1$, $f_0(t) = f(x+t) + f(x-t)/(2\pi)^{1/2}$), then it was proven that

$$S^\delta(R) = \lim_{\epsilon \to 0} cR^{(k/2)-\delta} \int_0^\infty e^{-\epsilon^2 t^2} f_0(t) J_{\delta+(k/2)}(tR)\, t^{(k/2)-\delta-1} dt, \tag{17}$$

where $c = 2^\delta \Gamma(\delta + 1)$, for all periodic functions $f(x_1, \cdots, x_k)$ of class L, and some more general almost periodic functions, for all real δ such that

$$\delta \geq 0, \tag{18}$$

and furthermore for all complex δ with

$$R(\delta) > 0. \tag{19}$$

Here $J_\mu(x)$ denotes the Bessel function of order μ. See [8]. We recall that only for

$$\delta > (k-1)/2 \tag{20}$$

is the integral

$$S^\delta(R) = cR^{(k/2)-\delta} \int_0^\infty f_0(t)\, t^{(k/2)-\delta-1} J_{\delta+(k/2)}(tR)\, dt \tag{21}$$

absolutely convergent, and relation (21) is the original version of a formula proved in [4].

We now state the object of the present paper. Chandrasekharan in [5, 6] has pointed out that (21), being a Hankel transform, has formally an inverse, namely

$$(22)\qquad f_0(t) = c_1 t^{\delta+2-(k/2)} \int_0^\infty R^{\delta+1-(k/2)} S^\delta(R) J_{\delta+(k/2)}(Rt)\,dR, \quad c_1 = 1/c.$$

Its precise validity poses a question [7], and an answer is, for instance, that it holds for almost all t provided that

$$(23)\qquad \delta > (k-1)/2.$$

We proceed to show that if we interpret (22) as

$$(24)\qquad f_0(t) = \lim_{\epsilon\to 0} c_1 t^{\delta+2-(k/2)} \int_0^\infty e^{-\epsilon^2 R^2} S^\delta(R) R^{\delta+1-(k/2)} J_{\delta+(k/2)}(Rt)\,dR,$$

then its validity is extensible to all values (18) and (19) of exponent δ. However, at present, some further assumption, different in nature, has to be added as regards the behavior of the function at the special point t at which the validity of (24) is desired. We shall see that (24) is valid at the point $t\,(0 < t < \infty)$ if and only if we have there

$$(25)\qquad (\pi)^{\frac{1}{2}} f_0(t) = \lim_{\epsilon\to 0} 1/\epsilon \int_{t/2}^{2t} f_0(\tau)\, e^{-(t-\tau/\epsilon)^2} d\tau,$$

and this is, in effect, a very mild continuity restriction on the function. Furthermore, we shall extend both (17) and (24) from the average $f_0(t)$ to its p-th mean defined by

$$(26)\qquad f_p(t) = 2\Gamma(p+k/2)/\{\Gamma(p)\Gamma(k/2)\, t^{k+2p-2}\} \int_0^t f_0(s)\,(t^2 - s^2)^{p-1} s^{k-1}\,ds,$$

in which case, (17) and (22) are to be replaced by

$$(27)\qquad S^\delta(R) = \lim_{\epsilon\to 0} c_2 R^{k+2p} \int_0^\infty t^{k+2p-1} e^{-\epsilon^2 t^2} f_p(t)\; V_{p+\delta+(k/2)}(tR)\,dt,$$

where

$$c_2 = 2^{\delta-p}\Gamma(\delta+1)\Gamma(k/2)/\Gamma(p+k/2),$$

and

$$(28)\qquad f_p(t) = \lim_{\epsilon\to 0} c_3 t^{2\delta+2} \int_0^\infty e^{-\epsilon^2 R^2} S^\delta(R) R^{2\delta+1} V_{p+\delta+(k/2)}(tR)\,dR, \quad c_3 = 1/c_2,$$

respectively, where $V_k(x) = J_k(x)/x^k$. See [5, 6]. They will again be valid for (18) and (19) in conjunction with

$$(29)\qquad R(p) > 0,$$

and

$$p \geq 0 \tag{30}$$

that is, for all possible combinations. We likewise notice that the pair of formulas (27) and (28) are Hankel inversions of each other, and our reasoning will, in effect, give criteria for the Gauss summability of such formulas, under suitable assumptions specifically arising in the theory of Fourier series.

2. The formula. We start from the formula

$$f_p(t) = c_3 t^{2\delta+2} \int_0^\infty S^\delta(\rho) \rho^{2\delta+1} \, V_{p+\delta+(k/2)}(t\rho) \, d\rho \tag{31}$$

for all possible combination of p and δ for which the right side converges absolutely. See [7]. Since for a fixed function $f(x_1, \cdots, x_k) \varepsilon L$ we have

$$S^\delta(\rho) = o(\rho^k) \tag{32}$$

uniformly in all x, the right side of (31) converges absolutely for

$$p - \delta - k/2 - \tfrac{1}{2} > 1, \tag{33}$$

and uniformly in every closed subset of this set (33) in the (p, δ) plane.

More precisely, we have, for $t > 0$,

$$\begin{aligned} t^{2\delta+2} \int_0^\infty | S^\delta(\rho) | \, \rho^{2\delta+1} \, | V_{p+\delta+(k/2)}(t\rho) | \, d\rho &= \int_0^{1/t} + \int_{1/t}^\infty \\ &= O(t^{2\delta+2} \int_0^{1/t} \rho^{k+2\delta+1} d\rho) + O(\int_{1/t}^\infty t^{\delta+3/2-p-k/2} / \rho^{p-\delta-k/2-\frac{1}{2}}) \, d\rho) \\ &= O(1/t^k). \end{aligned} \tag{34}$$

Now we introduce the 'approximate partial sum':

$$S^{\delta,\epsilon}(R) = c_2 R^{k+2p} \int_0^\infty e^{-\epsilon^2 t^2} t^{k+2p-1} f_p(t) \, V_{p+\delta+(k/2)}(tR) \, dt. \tag{35}$$

Since (33) implies that

$$p > 0, \tag{36}$$

on the basis of the estimate (34), we can substitute (31) in (35), and interchange the integrations with respect to t and ρ, thus obtaining

$$
\begin{aligned}
S^{\delta,\epsilon}(R) &= c_2 R^{k+2p-p-\delta-(k/2)} \int_0^\infty e^{-\epsilon^2 t^2} t^{k+2p-1-p-\delta-(k/2)} f_p(t) J_{p+\delta+(k/2)}(tR)\,dt \\
&= c_2 R^{(k/2)+p-\delta} \int_0^\infty e^{-\epsilon^2 t^2} t^{(k/2)+p-\delta-1} f_p(t) J_{p+\delta+(k/2)}(tR)\,dt \\
\text{(37)} \qquad &= c_2 R^{(k/2)+p-\delta} \int_0^\infty \int_0^\infty e^{-\epsilon^2 t^2} t J_{p+\delta+(k/2)}(t\rho) J_{p+\delta+(k/2)}(tR) S^\delta(\rho) \rho^{-(k/2)-p+\delta+1} d\rho\, dt \\
&= c_2 R^{(k/2)+p-\delta} \int_0^\infty S^\delta(\rho) \rho^{-(k/2)-p+\delta+1} d\rho \int_0^\infty e^{-\epsilon^2 t^2} t J_{p+\delta+(k/2)}(t\rho)\, J_{p+\delta+(k/2)}(tR)\,dt \\
&= c_2 R^{(k/2)+p-\delta} \int_0^\infty S^\delta(\rho) (2\epsilon^2)^{-1} \rho^{-(k/2)-p+\delta+1} e^{-(\rho^2+R^2)/4\epsilon^2} I_{p+\delta+(k/2)}(\rho R/\epsilon^2)\,d\rho
\end{aligned}
$$

on account of a formula in [8, p. 395]. Here I_ν denotes the Bessel function of imaginary argument.

Relation (37) is valid for the functions $S^\delta(R)$, $S^{\delta,\epsilon}(R)$ defined by (15) and (35), provided that p and δ are subject to the conditions

$$p > 0, \tag{38}$$

$$\delta > 0, \tag{39}$$

$$p - \delta - k/2 - 3/2 > 0, \tag{40}$$

of which (40) implies (38).

Now restriction (40) can be removed. In fact, expression (15) is analytic in δ for real part of $\delta > 0$; expression (26) is analytic in p for real part of $p > 0$. Furthermore, it follows from [8, p. 48(4)]

$$V_{p+\delta+(k/2)}(x) = (1/\Gamma(p+\delta+k/2+\tfrac{1}{2})\Gamma(\tfrac{1}{2})2^{p+\delta+(k/2)}) \int_{-1}^{+1} e^{ixt}(1-t^2)^{p+\delta+(k/2)-\frac{1}{2}} dt,$$

$R(p+\delta+(k/2)) > -\frac{1}{2}$, that we have the estimate

$$| V_{p+\delta+(k/2)}(x) | \leq c(A) e^{|x|}$$

for any compact subset A of the open set

$$R(p) > 0, \quad R(\delta) > 0,$$

of the four-dimensional Euclidean space

$$\delta = \delta_1 + i\delta_2, \qquad p = p_1 + ip_2.$$

Therefore, the right side of (35) converges absolutely and uniformly in every such set A, and (35), as a function of (p, δ) is complex analytic in A. Furthermore for such p and δ, the integral on the right of (37) converges in a similar manner, and is similarly analytic. Hence by analytic continuation, relation (37) also holds for (38) and (39) without (40). Finally, in all relations, it is admissible to let $p \downarrow 0$, and hence (37) also holds for

$$p \geqq 0, \tag{41}$$

$$\delta > 0; \tag{42}$$

actually it holds for

(43) the point-set defined by $R(p) > 0$ and $p = 0$,

(44) the point-set defined by $R(\delta) > 0$ and $\delta = 0$.

Having thus established (37), we now proceed to prove that

$$\lim_{\epsilon \to 0} S^{\delta, \epsilon}(R) = S^{\delta}(R). \tag{45}$$

Henceforward we shall consider only the case of real p and δ; the analysis can, however, be extended to the complex case with minor adjustments.

For any fixed R, we estimate the integral in (37) by splitting it into three parts:

$$R^{(k/2)+p-\delta} \int_0^\infty (S^\delta(\rho)/\epsilon^2) \rho^{-(k/2)-p+\delta+1} e^{-(\rho^2+R^2)/4\epsilon^2} I_{p+\delta+(k/2)}(\rho R/2\epsilon^2)\, d\rho$$

$$= \int_0^{R/2} + \int_{R/2}^{2R} + \int_{2R}^\infty. \tag{46}$$

The third integral in (46) is easy to estimate. For $0 < \epsilon \leqq \epsilon_0$ and $\rho \geqq 2R$, we have

$$| I_{p+\delta+(k/2)}(\rho R/2\epsilon^2) | = O(\epsilon e^{\rho R/2\epsilon^2}/(\rho R)^{1/2}) \tag{47}$$

and hence

$$\begin{aligned}
\Big|\int_{2R}^{\infty}\Big| &= O\big(R^{(k/2)+p-\delta}\int_{2R}^{\infty}\rho^{k-(k/2)-p+\delta+1}\epsilon^{-1}e^{-(\rho-R/2\epsilon)^2}(\rho R)^{-1/2}d\rho\big)\\
&= O\big(R^{(k/2)+p-\delta-\frac{1}{2}}\int_{2R}^{\infty}\rho^{(k/2)-p+\delta+\frac{1}{2}}e^{\rho^2/8\epsilon^2}\epsilon^{-1}d\rho\big)\\
&= O\big(R^{(k/2)+p-\delta-\frac{1}{2}}\epsilon^{(k/2)-p+\delta+\frac{1}{2}}\int_{2R/\epsilon}^{\infty}\rho^{(k/2)-p+\delta+\frac{1}{2}}e^{-\rho^2}d\rho\big)\\
&= o(1),\ \text{as}\ \epsilon\to 0.
\end{aligned}\tag{48}$$

To estimate the first integral in (46), we use $|S^{\delta}(R)| < c_4$, and consider the two subintervals

$$\rho R/\epsilon^2 \leq 1,\tag{49}$$

and

$$\rho R/\epsilon^2 > 1.\tag{50}$$

In (49) we have

$$I_{p+\delta+(k/2)}(\rho R/\epsilon^2) = O(\rho R/\epsilon^2)^{p+\delta+(k/2)}$$

so that

$$\begin{aligned}
\Big|\int_{0}^{\epsilon^2/R}\Big| &= O\big(R^{(k/2)+p-\delta}\int_{0}^{\epsilon^2/R}(e^{-R^2/4\epsilon^2}\rho^{2\delta+1}R^{p+\delta+(k/2)}d\rho/\epsilon^{2+2p+2\delta+k}\big)\\
&= O\big(R^{2p+k}e^{-R^2/4\epsilon^3}/\epsilon^{2+2p+2\delta+k}\int_{0}^{\epsilon^2/R}\rho^{2\delta+1}d\rho\big)\\
&= o(1),\ \text{as}\ \epsilon\to 0.
\end{aligned}\tag{51}$$

In (50) we again have estimate (47) and thus

$$\Big|\int_{\epsilon^2/R}^{R/2}\Big| = O\big(R^{(k/2)+p-\delta-\frac{1}{2}}\int_{\epsilon^2/R}^{R/2}\rho^{-(k/2)-p+\delta+\frac{1}{2}}\epsilon^{-1}e^{-(\rho-R/2\epsilon)^2}d\rho\big)$$

but here

$$e^{-(\rho-R/2\epsilon)^2} = O(e^{-R^2/4\epsilon^2})$$

and thus

$$\Big|\int_{\epsilon^2/R}^{R/2}\Big| = o(1),\ \text{as}\ \epsilon\to 0.\tag{52}$$

From (48), (51) and (52), we obtain the result

$$\begin{aligned}
&S^{\delta,\epsilon}(R)R^{-(k/2)-p+\delta}\\
&\sim c_2\int_{R/2}^{2R}S^{\delta}(\rho)\rho^{-(k/2)-p+\delta+1}e^{-(\rho^2+R^2)/4\epsilon^2}I_{p+\delta+(k/2)}(\rho R/2\epsilon^2)(1/2\epsilon^2)\,d\rho.
\end{aligned}$$

Now, in $R/2 \leq \rho \leq 2R$, we use

$$I_\nu(\rho R/2\epsilon^2) = (c_5\epsilon/(\rho R)^{\frac{1}{2}})e^{\rho R/2\epsilon^2} + O((\epsilon^2/\rho R)e^{\rho R/2\epsilon^2}). \tag{53}$$

The error-term in (53) gives rise to

$$O(1)1/R\int_{R/2}^{2R}\rho^{(k/2)-p+\delta}e^{-(\rho-R/2\epsilon)^2}d\rho,$$

but for fixed R, this tends to zero as $\epsilon \rightarrow 0$, and thus can be neglected. Hence we are left with

$$S^{\delta,\epsilon}(R)R^{-(k/2)-p+\delta+\frac{1}{2}} \sim \int_{R/2}^{2R}S^\delta(\rho)\rho^{-(k/2)-p+\delta+\frac{1}{2}}e^{-(\rho-R/2\epsilon)^2}d\rho$$

If we put $\gamma = -(k/2) - p + \delta + \frac{1}{2}$ we have

$$[S^{\delta,\epsilon}(R) - S^\delta(R)] \sim c_2\int_{R/2}^{2R}[S^\delta(\rho) - S^\delta(R)](\rho/R)^\gamma e^{-(\rho-R/2\epsilon^2)^2}d\rho.$$

But

$$(\rho/R)^\gamma - 1 = O(|\rho - R|)$$

and hence

$$S^{\delta,\epsilon}(R) - S^\delta(R) \sim c_2\int_{R/2}^{2R}[S^\delta(\rho) - S^\delta(R)]\epsilon^{-1}e^{-(\rho-R/2\epsilon)^2}d\rho.$$

Since $S^\delta(R)$ is a continuous function of R for $R(\delta) > 0$, this implies (45). For $\delta = 0$, $S^\delta(R) = (S^\delta(R+0) + S^\delta(R-0))/2$.

3. The inverse formula.

We want to establish the relation

$$f_p(t) = \lim_{\epsilon\rightarrow 0} c_3 t^{2\delta+2}\int_0^\infty e^{-\epsilon^2R^2}S^\delta(R)R^{2\delta+1}V_{p+\delta+(k/2)}(tR)\,d\mathrm{R} \tag{54}$$

for certain values of t, for all $R(p) \geqq 0$, $R(\delta) > 0$. For

$$\delta > p + (k-1)/2 \tag{55}$$

we have

$$S^\delta(R) = c_2R^{k+2p}\int_0^\infty \tau^{k+2p-1}f_p(\tau)V_{p+\delta+(k/2)}(\tau R)\,d\tau. \tag{56}$$

See [5, 6]. If we substitue this in the right side of (53) we obtain

$$
\begin{aligned}
f_p(t) &= c_3 t^{2\delta+2} \int_0^\infty \int_0^\infty e^{-\epsilon^2 R^2} R^{2\delta+1+k+2p} \tau^{k+2p-1} f_p(\tau)\, V_{p+\delta+(k/2)}(tR) \ \ V_{p+\delta+(k/2)}(\tau R)\, dR\, d\tau \\
&= c_3 t^{2\delta+2} \int_0^\infty \int_0^\infty e^{-\epsilon^2 R^2} R t^{-p-\delta-(k/2)} \tau^{(k/2)+p-\delta-1} f_p(\tau)\, J_{p+\delta+(k/2)}(tR) \ \ J_{p+\delta+(k/2)}(\tau R)\, dR\, d\tau \\
&= c_3 t^{\delta-p-(k/2)+2} \int_0^\infty \tau^{(k/2)+p-\delta-1} f_p(\tau)\, d\tau \int_0^\infty e^{-\epsilon^2 R^2} R J_{p+\delta+(k/2)}(tR) \ \ J_{p+\delta+(k/2)}(\tau R)\, dR \\
&= c_3 t^{\delta-p-(k/2)+2} \int_0^\infty \tau^{(k/2)+p-\delta-1} f_p(\tau)\, e^{-(t^2+\tau^2)/4\epsilon^2} I_{p+\delta+(k/2)}(t\tau/2\epsilon^2)\, d\tau,
\end{aligned}
$$

and by a similar argument as in section 2, this is equiconvergent with

$$
c_3/\epsilon \int_{t/2}^{2t} e^{-(t-\tau)/2\epsilon^2} f_p(\tau)\, d\tau.
$$

PRINCETON UNIVERSITY
INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

1. S. Bochner, "Properties of Fourier series of almost periodic functions," *Proceedings of the London Mathematical Society* (2), vol. 26 (1927), pp. 433-452.
2. ———, "Über Fourierreihen von fastperiodischen Funktionen," *Sitzungsberichte der Berliner Mathematische Gesellschaft*, vol. 26 (1927), pp. 49-70.
3. ———, "On spherical partial sums of multiple Fourier series," To appear in *Revista di Ciencias*.
4. ———, "Summation of multiple Fourier series by spherical means," *Transactions of the American Mathematical Society*, vol. 40 (1936), pp. 175-207.
5. K. Chandrasekharan, "On multiple Fourier series," *Proceedings of the Indian Academy of Sciences*, vol. XXIV (1946), pp. 229-232.
6. ———, "On the summation of multiple Fourier series I," *Proceedings of the London Mathematical Society* (2), vol. 50 (1948), pp. 210-222.
7. ———, "On Fourier series in several variables," *Annals of Mathematics*, vol. 49 (1948), pp. 991-1007.
8. G. N. Watson, *A treatise on Bessel functions*, Cambridge, 1922.

NOTES ON FOURIER EXPANSIONS III.*

(*Fourier Stieltjes Series*)

By S. MINAKSHISUNDARAM.

1. If $F(x)$ is a function of bounded variation defined in the interval $0 \leq x \leq 2\pi$, and if

$$dF(x) \sim \tfrac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \tag{1}$$

then we are aware of the following results on the summability of the Fourier Stieltjes Series on the right of (1):

i) *Almost everywhere the Fourier Stieltjes Series is summable* (C, α), *for every* $\alpha > 0$.

ii) *At a point where* $F(x)$ *has a derivative the series is summable* $(C, 1+\alpha)$ *for every* $\alpha > 0$.

A generalization of this result to k-dimensional space will read as follows, if we make use of spherical summation:

Let $F(e)$ *be an additive function of bounded variation defined on Borel sets in the interval* $(0 \leq x_1, x_2, \cdots x_k \leq 2\pi)$ *and let*

$$\begin{aligned} dF(e) &\sim \sum C_{n_1\ldots n_k} e^{i(n_1x_1+\ldots+n_kx_k)} \\ C_{n_1\ldots n_k} &= (1/(2\pi)^k) \int_0^{2\pi} \cdots \int_0^{2\pi} e^{-i(n_1x_1+\ldots+n_kx_k)}\, dF(e). \end{aligned} \tag{2}$$

Then

i) *the series on the right of* (2) *called the Multiple Fourier Stieltjes Series, is almost everywhere summable* $(\nu^2, (k-1/2)+\epsilon)$ *for every* $\epsilon > 0$ *i. e.*

$$\lim_{R\to\infty} \sum_{\nu^2 \leq R^2} (1 - (\nu^2/R^2))^{(k-1/2)+\epsilon} C_{n_1\ldots n_k} e^{i(n_1x_1+\ldots+n_kx_k)}, \quad \nu^2 = n_1{}^2 + \cdots + n_k{}^2$$

exists almost everywhere;

ii) *at a point where the symmetric derivative of* $F(e)$ *exists the series on the right of* (2) *is summable* $(\nu^2, (k+1/2)+\epsilon)$ *for every* $\epsilon > 0$ *to the symmetric derivative of* $F(e)$ *i. e.*

---

* Received June 12, 1947.

$$\lim_{R\to\infty} \sum_{\nu^2\leq R^2} (1-(\nu^2/R^2))^{(k+1/2)+\epsilon} C_{n_1\ldots n_k} e^{i(n_1x_1+\cdots+n_kx_k)} = D_{sym.} F(e).$$

The main aim of this note is to prove these results. We prove them for the two dimensional case only for the sake of simplicity, but the method is quite general and can be easily applied to several dimensions. Incidentally we give a new proof of Bochner's result on spherical summation of multiple Fourier series [1].

2. In the Euclidean x, y plane, let (e) denote Borel sets contained in the fundamental square $(-\frac{1}{2}\leq x, y\leq\frac{1}{2})$ and let $e_{\xi,\eta}$ denote the set of points $(\xi+x, \eta+y)$, $(x,y)\, \varepsilon\, e$. Further we denote by s^r the interior of the circle $x^2+y^2=r^2$ and $s^r{}_{\xi,\eta}$ the circular region with center (ξ,η) and radius r. We further denote by $F(e)$ an additive function of Borel sets (e) defined in $-\frac{1}{2}\leq x, y\leq\frac{1}{2}$, which is of bounded variation in this square. We note that this can be extended throughout the plane by the conditions

$$F(e_{k,1}) = F(e),\ k, 1 \text{ being integers.}$$

If at a point

$$\lim_{r=0} (F(s^r{}_{\xi,\eta})/\pi r^2) = f(\xi,\eta)$$

exists, then we say that $F(e)$ has the symmetric derivative $f(\xi,\eta)$ at the point (ξ,η). It will be convenient to write

$$F(s^r{}_{\xi,\eta}) = \phi(\xi,\eta;r) = \phi(r).$$

It should be noted that, for every fixed ξ, η, $\phi(r)$ is of bounded variation in r. Finally we denote by C a positive constant independent of all the variables involved; though it may occur in different places, it does not necessarily take the same value.

3. In addition to a few important properties of Bessel functions, we require a lemma due to Hardy and Landau and a transformation formula for theta functions, which will be made use of for obtaining a formula for spherical averages of the double Fourier series. The transformation formula is

$$(3)\qquad g(u) = g(u,\delta) = \sum_{m=-\infty}^{\infty} e^{-(m-n)^2/\delta^2} = \delta\pi^{\frac{1}{2}} \sum_{-\infty}^{\infty} e^{-m^2\pi^2\delta^2+2\pi miu}$$

and the lemma is

Lemma 1. *If K is the circle $u^2+v^2\leq R^2$, $\psi(u,v)$ is any continuous function, then*

$$(4) \qquad J_\delta = 1/\pi\delta^2 \int_K \int g(u)g(v)\psi(u,v)\,dudv \to \textstyle\sum' \psi(\mu,\nu), \qquad \delta \to 0$$

where summation extends over the lattice points of K and the dash implies that the lattice points on the boundary are affected by a weight $\frac{1}{2}$.

For the proof cf. G. H. Hardy [2].

In this lemma let us take

$$\psi(u,v) = (R^2 - u^2 - v^2)^\alpha e^{-2\pi i(ux+vy)}, \qquad \alpha > 0$$

we have then

$$(5) \qquad \sum_{\mu^2+\nu^2 \leq R^2} (R^2 - \mu^2 - \nu^2)^\alpha e^{-2\pi i(\mu x+\nu y)}$$
$$= \lim_{\delta\to 0} \tfrac{1}{2} \int_K \int g(u)g(v)(R^2 - u^2 - v^2) e^{-2\pi i(\mu x+\nu y)}\,dudv.$$

On the right of (5) we make use of the transformation formula (3) and obtain

$$\sum_{\mu^2+\nu^2 \leq R^2} (R^2 - \mu^2 - \nu^2)^\alpha e^{-2\pi i(\mu x+\nu y)}$$
$$= \lim_{\delta\to 0} \int_K \int (R^2 - u^2 - v^2)^\alpha \sum e^{-\pi^2\delta^2(m^2+n^2)+2\pi i((m-x)u+(n-y)v)}\,dudv$$
$$= \lim_{\delta\to 0} \sum e^{-\pi^2\delta^2(m^2+n^2)} \int_K \int (R^2 - u^2 - v^2)^\alpha e^{2\pi i((m-x)u+(n-y)v)}\,dudv$$
$$(6) \qquad = 2\pi \lim_{\delta\to 0} \sum e^{-\pi^2\delta^2(m^2+n^2)} \int_0^R (R^2 - t^2)^\alpha J_0(2\pi t\{(m-x)^2 + (n-y)^2\}^{\frac{1}{2}} t\,dt$$
$$= \frac{\Gamma(\alpha+1)}{\pi^\alpha} R^{1+\alpha} \lim_{\delta\to 0} \sum e^{-\pi^2\delta^2(m^2+n^2)} \frac{J_{1+\alpha}(2\pi R\{(m-x)^2 + (n-y)^2\}^{\frac{1}{2}}}{\{(m-x)^2 + (n-y)^2\}^{(1+\alpha)/2}}$$
$$= \frac{\Gamma(\alpha+1)}{\pi^\alpha} R^{\alpha+1} \sum \frac{J_{1+\alpha}(2\pi R\{(m-x)^2 + (n-y)^2\}^{\frac{1}{2}}}{\{(m-x)^2 + (n-y)^2\}^{(1+\alpha)/2}}$$

the series in the last line being absolutely convergent for $\alpha > \frac{1}{2}$. If x and y lie in the interval $(-\frac{1}{2}, \frac{1}{2})$ we observe that

$$(m - \tfrac{1}{2})^2 + (n - \tfrac{1}{2})^2 \leq (m-x)^2 + (n-y)^2 \leq (m + \tfrac{1}{2})^2 + (n + \tfrac{1}{2})^2,$$

so that

$$1/(m-x)^2 + (n-y)^2 = O(1/m^2 + n^2) \text{ if } m^2 + n^2 > 0,$$

and hence

$$\sum_{m^2+n^2>0} J_{1+\alpha}(2\pi R\{(m-x)^2 + (n-y)\}^{\frac{1}{2}}) / \{(m-x)^2 + (n-y)^2\}^{(1+\alpha)/2}$$
$$= O\{1/R^{\frac{1}{2}} \textstyle\sum 1/(m^2+n^2)^{2\alpha+3/4}\}$$
$$= O(R^{-\frac{1}{2}}).$$

So we can write (6) in the form

$$(7) \qquad \sum (R^2 - \mu^2 - \nu^2)^{\alpha} e^{-2\pi i(\mu x + \nu y)} = (\Gamma(\alpha+1)/\pi^{a}) R^{1+a} (J_{1+a}(2\pi Rr)/r^{1+a}) + O(R^{a+\frac{1}{2}})$$

uniformly in x and y, if $\alpha > \frac{1}{2}$ where $x^2 + y^2 = r^2$.

4. Now let $F(e)$ be an additive function of bounded variation; then from (7)

$$(8) \qquad \begin{aligned} &\sum (1 - (\mu^2 + \nu^2/R^2))^{a} \int\int e^{-2\pi i(\mu x + \nu y)} dF(e_{\xi,\eta}) \\ &= (\Gamma(\alpha+1) R^{1-a}/\pi^{a}) \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} (J_{1+a}(2\pi Rr)/r^{1+a})\, dF(e_{\xi,\eta}) \\ &\quad + O(R^{-a+\frac{1}{2}}). \end{aligned}$$

Again, if τ is a positive number, to be fixed later,

$$\int\int_{x^2+y^2 \geq \tau^2} (J_{1+a}(2\pi Rr)/r^{1+a}) dF(e_{\xi,\eta}) = O(R^{-\frac{1}{2}}).$$

so that (8) can be rewritten in the form

$$(9) \qquad \begin{aligned} &\sum (1 - (\mu^2 + \nu^2/R^2)^{a} c_{\mu\nu} e^{2\pi i(\mu x + \nu y)} \\ &= (\Gamma(\alpha+1)/(\pi R)^{a-1}) \int_0^{\tau} (J_{1+a}(2\pi Rr)/r^{1+a}) d\phi(r) \\ &\quad + O(R^{-a+\frac{1}{2}}). \end{aligned}$$

Now consider the integral on the right of (9) viz.

$$\begin{aligned} 1/R^{a-1} \int_0^{\tau} (J_{1+a}(2\pi Rr)/r^{1+a}) d\phi(r) &= 1/R^{a-1} \left[\int_0^{1/R} + \int_{1/R}^{\tau} \right] \\ &= J_1 + J_2 \end{aligned}$$

say. If we set

$$\int_0^{r} | d\phi(r) | = \Phi(r)$$

then

$$\begin{aligned} | J_1 | &= | 1/R^{a-1} \int_0^{1/R} (J_{1+a}(2\pi Rr)/r^{1+a}) d\phi(r) | \\ &\leq CR^2 \int_0^{1/R} | d\phi(r) | = CR^2 \Phi(1/R) \end{aligned}$$

$$| J_2 | = | 1/R^{a-1} \int_{1/R}^{\tau} (J_{1+a}(2\pi Rr)/r^{1+a}) d\phi(r) |$$

$$
\begin{aligned}
&\leq C/R^{\alpha-\frac{1}{2}} \int_{1/R}^{\tau} (1/r^{\alpha+3/2}) \, | \, d\phi(r) \, | \\
&= C/R^{\alpha-\frac{1}{2}} [\Phi(r)/r^{\alpha+3/2}]_{1/R}^{\tau} + C/R^{\alpha-\frac{1}{2}} \int_{1/R}^{\tau} (\Phi(r)/r^{\alpha+5/2}) \, dr \\
&= (C\Phi(\tau)/R^{\alpha-\frac{1}{2}}\tau^{\alpha+3/2}) - CR^2\Phi(1/R) + C/R^{\alpha-\frac{1}{2}} \int_{1/R}^{\tau} (\Phi(r)/r^{\alpha+5/2}) \, dr.
\end{aligned}
$$

If $\Phi(r) = o(r^2)$ then given $\epsilon > 0$ we can choose τ so that for $r \leq \tau$

$$| \, \Phi(r) \, | \leq \epsilon r^2$$

and then

$$| \, J_2 \, | < C\epsilon$$

$$| \, J_2 \, | < C\epsilon + C\epsilon + (C\epsilon/R^{\alpha-\frac{1}{2}}) \int_{1/R}^{\tau} (dr/r^{\alpha+\frac{1}{2}}) < C\epsilon.$$

Therefore we have

$$J_1 + J_2 = o(1) \qquad \text{as } R \to \infty.$$

If $\Phi(r) = r^2(s + o(1))$ then we can show similarly that $J_1 + J_2 = s + o(1)$ as $R \to \infty$. *Thus at all points where* $\Phi(r) = r^2[s + o(1)]$ *and therefore almost everywhere the series is summable* (ν^2, α) $\alpha > \frac{1}{2}$.

5. We shall now show that *at a point where* $\phi(r) = o(r^2)$ [or $\phi(r) = r^2(s + o(1))$] *the Fourier Stieltjes series is summable* (ν^2, α) *for* $\alpha > 3/2$. We have only to show that

$$\text{(9)} \qquad 1/R^{\alpha-1} \int_{R}^{\tau} (J_{1+\alpha}(2\pi Rr)/r^{1+\alpha}) \, d\phi(r) = o(1) \text{ as } R \to \infty, \text{ for } \alpha > 3/2.$$

On integrating the integral on the left by parts we obtain

$$
\begin{aligned}
&(J_{1+\alpha}(2\pi Rr)/r^{1+\alpha})\phi(r) \Big]_0^{\tau} - CR \int_0^{\tau} (J_{2+\alpha}(2\pi Rr)/r^{1+\alpha}) \phi(r) \, dr \\
&\quad = (J_{1+\alpha}(2\pi R\tau)/\tau^{1+\alpha}) \phi(\tau) - CR \int_0^{\tau} (J_{2+\alpha}(2\pi Rr)/r^{1+\alpha}) \phi(r) \, dr \\
&\quad = J_1 + J_2
\end{aligned}
$$

say. Let

$$| \, \phi(r) \, | \leq \epsilon r^2 \text{ for } r \leq \tau.$$

Then

$$| \, J_1 \, | < \epsilon/R^{\frac{1}{2}}\tau^{\alpha-\frac{1}{2}}$$

$$| \, J_2 \, | \leq \epsilon \int_0^{\tau R} (| \, J_{2+\alpha}(2\pi y)/y^{\alpha-1}) \, dy < C\epsilon \qquad \text{if } \alpha > 3/2.$$

Hence (9) follows. If $\phi(r) = r^2(s + o(1))$, the right-side of (9) will be $s + o(1)$. It should be observed that $\phi(r^2) = r^2(s + O(1))$ means that $F(e)$ has the symmetric derivative s/π at the point.

6. We shall now use the formula (6) to obtain the fundamental formula of Bochner: Let $f(\xi, \eta)$ be an integrable function periodic in each of the variables with period 1 and let us write

$$g(x, y) = f(\xi + x, \eta + y)$$

so that the behaviour of $f(\xi, \eta)$ in the neighborhood of (ξ, η) is the same as the behaviour of $g(x, y)$ near the origin. Further $g(x, y)$ is also periodic and if

$$f(\xi, \eta) \sim \sum C_{\mu\nu} e^{2\pi i(\mu\xi + \nu\eta)}$$

$$C_{\mu\nu} = \int\int f(\xi, \eta)\, e^{-2\pi i(\mu\xi+\nu\eta)}\, d\xi d\eta$$

then

$$\begin{aligned} g(x, y) &\sim \sum d_{\mu\nu} e^{2\pi i(\mu x + \nu y)} \\ &= \sum c_{\mu\nu} e^{2\pi i(\mu(x+\xi) + \nu(y+\eta))}. \end{aligned}$$

Further let

$$f_{\xi,\eta}(t) = \tfrac{1}{2\pi}\int_0^{2\pi} f(\xi + t\cos\theta, \eta + t\sin\theta)\, d\theta = \tfrac{1}{2\pi}\int_0^{2\pi} g(t\cos\theta, t\sin\theta)\, d\theta.$$

Now multiply both sides of (6) by $g(x, y)$ and integrate over the region $0 \leq x, y \leq 1$. The left side will give us

$$\sum_{\mu^2+\nu^2 \leq R^2} (R^2 - \mu^2 - \nu^2)^a d_{\mu\nu} = \sum (R^2 - \mu^2 - \nu^2)^a c_{\mu\nu} e^{2\pi i(\mu\xi+\nu\eta)}$$

while the right side leads to

$$\begin{aligned} &\frac{\Gamma(\alpha + 1)R^{a+1}}{\pi^a} \sum \int_0^1 \int_0^1 \frac{J_{1+a}(2\pi R\{(m-x)^2 + (n-y)^2\}^{\frac{1}{2}})}{\{(m-x)^2 + (n-y)^2\}^{(1+a)/2}}\, g(x, y)\, dxdy \\ &\quad = \frac{\Gamma(\alpha + 1)R^{1+a}}{\pi^a} \sum \int_{R_m} \int \frac{J_{1+a}(2\pi Rr)}{r^{1+a}}\, g(m-u, n-u)\, dudv, \end{aligned}$$

or, if $u^2 + v^2 = r^2$,

$$\begin{aligned} &= \frac{\Gamma(\alpha + 1)R^{a+1}}{\pi^a} \int_0^\infty \int_0^\infty \frac{J_{1+a}(2\pi Rr)}{r^{1+a}}\, g(u, v)\, dudv \\ &= \frac{2\Gamma(\alpha + 1)R^{a+1}}{\pi^{a-1}} \int_0^\infty \frac{J_{1+a}(2\pi Rt)}{t^a}\, f_{\xi,\eta}(t)\, dt, \end{aligned}$$

where R_m is the rectangle $m - 1 \leq x, y \leq m$.

Hence we have the formula of Bochner [1]

$$(10) \quad \sum (1 - (\mu^2 + \nu^2/R^2))^a c_{\mu\nu} e^{2\pi i(\mu\xi+\nu\eta)} = \frac{2\Gamma(\alpha+1)}{(\pi R)^{a-1}} \int_0^\infty \frac{J_{1+a}(2\pi Rt)}{t^a} f_{\xi,\eta}(t)$$

If, however, we multiply both sides of (7) by $g(x,y)$ and integrate over the region $-\frac{1}{2} \leq x, y \leq \frac{1}{2}$ we have

$$(11) \quad \sum (R^2 - \mu^2 - \nu^2)^a c_{\mu\nu} e^{2\pi i(\mu\xi+\eta\nu)} = \frac{\Gamma(\alpha+1)R^{1+a}}{\pi^a} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{J_{1+a}(2\pi Rr)}{r^{1+a}} g(x,y)\,dxdy + O(R^{-\frac{1}{2}})$$

and if we observe that for any $\tau > 0$

$$\iint\limits_{x^2+y^2 \geqq \tau} \frac{J_{1+a}(2\pi Rr)}{r^{1+a}} g(x,y)\,dxdy = O(R^{a+\frac{1}{2}}), \qquad \alpha > \tfrac{1}{2}$$

then (11) leads to

$$(12) \quad \sum (R^2 - \mu^2 - \nu^2)^a c_{\mu\nu} e^{2\pi i(\mu\xi+\nu\eta)} = \frac{2\Gamma(\alpha+1)R^{1+a}}{\pi^{a-1}} \int_0^\tau \frac{J_{1+a}(2\pi Rr)}{r^a} f_{\xi,\eta}(r)\,dr + O(R^{a+\frac{1}{2}})$$

or what is the same thing

$$(13) \quad \sum (1 - (\mu^2 + \nu^2/R^2))^a c_{\mu\nu} e^{2\pi i(\mu\xi+\nu\eta)} = \frac{2\Gamma(\alpha+1)R^{1-a}}{\pi^{a-1}} \int_0^\tau \frac{J_{1+a}(2\pi Rt)}{t^a} f_{\xi,\eta}(t)\,dt + O(1/R^{a-\frac{1}{2}}).$$

which proves the local property for summability of order $\alpha > \frac{1}{2}$.

INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

REFERENCES.

1. S. Bochner, "Summation of multiple Fourier series by spherical means," *Transactions of the American Mathematical Society*, vol. 40 (1936), pp. 175-207.
2. G. H. Hardy, "The lattice points of a circle," *Proceedings of the Royal Society of London*, vol. 107 (1925), pp. 623-635.

ON A LIAPOUNOFF CRITERION OF STABILITY.*

By Göran Borg.

The following statements will be proved in this paper:

If, in the differential equation

$$y'' + \phi(x)y = 0, \tag{L}$$

the coefficient function $\phi(x)$ *is continuous, of period* π, $\phi(x+\pi) = \phi(x)$, *and satisfies*

$$\int_0^\pi \phi(x)\,dx \geq 0 \text{ and } \phi(x) \not\equiv 0 \tag{1}$$

and

$$\int_0^\pi |\phi(x)|\,dx \leq 4/\pi, \tag{2}$$

then (L) *is stable. These conditions are the "best possible" in the following sense: There are unstable differential equations* (L) *for which*

$$\int_0^\pi \phi(x)\,dx \geq -\epsilon \text{ and } \phi(x) \not\equiv 0 \tag{1*}$$

and (2) *hold, or for which* (1) *and*

$$\int_0^\pi |\phi(x)|\,dx \leq \epsilon + 4/\pi \tag{2*}$$

hold, where $\epsilon > 0$ *is arbitrary.*

This result completes a criterion of Liapounoff for stability by replacing the Liapounoff condition $\phi(x) \geq 0$ and $\phi(x) \not\equiv 0$ by (1).

Proof. In order to prove the first assertion above, suppose, if possible, that (L) is unstable. Then there exists a real solution $y = y(x)$ of (1) satisfying

$$y(x+\pi) = sy(x) \text{ and } y(x) \not\equiv 0, \tag{3}$$

where $s \neq 0$ is a real number. Clearly, $y(x)$ has either no zeros or an infinity of zeros. In the latter case, the distance between adjacent zeros does not

---

* Received April 21, 1948.

exceed π. It will be shown in (**A**) and (**B**) that assumptions (1) and (2) contradict both of these alternatives.

(**A**) Suppose that $y(x)$ does not vanish, then by (L)

$$\int_0^\pi y''y^{-1}dx + \int_0^\pi \phi dx = 0.$$

An integration by parts gives

$$y'y^{-1} \Big|_0^\pi + \int_0^\pi y'^2y^{-2}dx + \int_0^\pi \phi dx = 0.$$

Since $y'(\pi)y^{-1}(\pi) = ss^{-1}y'(0)y^{-1}(0)$, it follows that either

$$\int_0^\pi \phi dx < 0 \text{ or } y' \equiv 0.$$

These two possibilities contradict the assumptions in (1).

(**B**) Suppose that $y(x)$ possesses zeros. It can then be supposed that $0 < b - a \leq \pi$, that $y(a) = y(b) = 0$, and that $y(x) > 0$ on (a, b). It will be shown that these facts, in addition to the existence and continuity of the derivatives y' and y'', imply[1]

$$\int_a^b | y''y^{-1} | dx > 4/(b-a). \tag{4}$$

In fact, it is clear that

$$\int_a^b | y''y^{-1} | dx > (y_{\max})^{-1} \max_{a \leqq \xi < \eta \leqq b} | y'(\xi) - y'(\eta) | . \tag{5}$$

But, if $y_{\max} = y(a + l_1) = y(b - l_2)$, where $l_1 + l_2 = b - a$, then it follows from Rolle's theorem that

$$y'(\xi) = l_1^{-1}y_{\max} \text{ and } -y'(\eta) = l_2^{-1}y_{\max}$$

for some choice of ξ and η, where $a < \xi < a + l_1 = b - l_2 < \eta < b$. Hence, from (5),

$$\int_a^b | y''y^{-1} | dx > l_1^{-1} + l_2^{-1} = (l_1 + l_2)(l_1l_2) \geq 4/(b-a),$$

in view of the arithmetic-geometric inequality. This proves (4).

---

[1] This inequality is due to Beurling. In my paper, "Ueber die Stabilität gewisser Klassen von linearen Differentialgleichungen," *Arkiv för Matematik, Astronomi och Fysik*, vol. 31, no. 1 (1945), pp. 1-31, the inequality is generalized and used for the proofs of other, in most respects, more far-reaching stability criteria than the one above. However, the one above is not contained explicitly in these results.

In view of $b-a\leq\pi$, (4) and (L) contradict (2) and, therefore, complete the proof of the first assertion of the italicized statement.

(C) It will now be shown that (1*) and (2) do not imply the stability of (L). To this end, let $\frac{1}{2}\pi>\eta>0$ and $k>0$. Let $y=y(x)\geq 1$ be a function which has a period π, possesses a continuous second derivative, is identically 1 on the intervals $(0,\frac{1}{2}\pi-\eta)$ and $(\frac{1}{2}\pi+\eta,\pi)$, and has a derivative satisfying $|y'|<k$ and $y'\not\equiv 0$. Define $\phi(x)$ to be $-y''(x)y^{-1}(x)$; so that $\phi(x)$ is continuous and of period π. Also

$$\int_0^{\pi}\phi(x)\,dx=-\int_{\frac{1}{2}\pi-\eta}^{\frac{1}{2}\pi+\eta}y''y^{-1}dx=-\int_{\frac{1}{2}\pi-\eta}^{\frac{1}{2}\pi+\eta}y'^2y^{-2}dx\geq-2k^2\eta.$$

Hence, if $\epsilon>0$ is given, (1*) is satisfied if k and η are chosen sufficiently small. Furthermore, (2) is satisfied whenever k is sufficiently small.

It remains to show that (L) is of unstable type. Since the function $y(x)$ is a periodic solution of (L), the equation (L) can be stable only if all solutions of (L) are periodic. In this case, the solution $y=z(x)$ of (L) satisfying the initial condition $z(0)=0$ (and $z'(0)=1$) has an infinity of zeros. This leads to a contradiction, since the Sturm separation theorem implies that a non-trivial solution of (L) has at most one zero. Consequently, (L) is unstable.

(D) It will now be shown that (1) and (2*) do not imply the stability of (L). To this end, let $\frac{1}{2}\pi>\eta>0$ and let $y=y(x)$ be an odd function of period 2π, possessing a non-positive continuous second derivative on $0\leq x\leq\pi$, and, in addition,

$$y(x)=\begin{cases}x \text{ for } 0\leq x\leq\frac{1}{2}\pi-\eta,\\ \pi-x \text{ for } \frac{1}{2}\pi+\eta\leq x\leq\pi.\end{cases}$$

Define $\phi(x)$ to be $-y''y^{-1}$ or 0 according as $y\neq 0$ or $y=0$. Then $\phi(x)$ is a continuous function of period π. Also, $\phi(x)\geqq 0$, so that (1) is satisfied. As above,

$$\int_0^{\pi}|\phi(x)|\,dx=-\int_{\frac{1}{2}\pi-\eta}^{\frac{1}{2}\pi+\eta}y''y^{-1}dx;$$

but the last expression is not greater than

$$(\tfrac{1}{2}\pi-\eta)^{-1}(y'(\tfrac{1}{2}\pi-\eta)-y'(\tfrac{1}{2}\pi+\eta))=4/(\pi-2\eta),$$

so that (2*) is satisfied if η is sufficiently small.

It remains to show that (L) is of unstable type. Since the function $y=y(x)$ is a half-periodic solution of (L), that is, $y(x+\pi)=-y(x)$, the equation (L) can be stable only if all solutions of (L) are half-periodic. Suppose (L) is stable. Consider the solution $y = z(x)$ of (L) satisfying the initial condition $z(a) = 0$ and $z'(a) = m > 0$, where $a = \frac{1}{2}\pi - \eta - \delta$ and $0 < \delta < \frac{1}{2}\pi - \eta$. Then $z(a + \pi) = 0$ and $z'(a + \pi) = -m$. Since $\phi(x) = 0$ for $0 \leq x \leq \frac{1}{2}\pi - \eta$ and $\frac{1}{2}\pi - \eta \leq x \leq 3\pi/2 - \eta$, it follows from (L) that $z(x)$ is linear on these intervals, so that

$$z(\tfrac{1}{2}\pi + \eta) = m(a + \pi - \tfrac{1}{2}\pi - \eta) = m(\pi - 2\eta - \delta).$$

Also, since $\phi(x) \geq 0$, it is seen from (L) that the solution $z(x)$ is concave downward where $z(x) > 0$. Hence, on the interval $a < x < a + \pi$, the graph of $y = z(x)$ does not cross the line $y = m(x - a)$. In particular,

$$z(\tfrac{1}{2}\pi + \eta) \leq m(2\eta + \delta).$$

If the arbitrary positive numbers η, δ are suitably chosen, the last two formula lines lead to a contradiction. This shows that (L) is unstable and completes the proof of the italicized statement.

The statement in **(D)** was demonstrated in a different way by van Kampen and Wintner.[2]

MATEMATISKA INSTITUTIONEN,
UPPSALA, SWEDEN.

---

[2] E. R. van Kampen and A. Wintner, "On an absolute constant in the theory of variational stability," *American Journal of Mathematics*, vol. 59 (1937), pp. 270-274.

ON THE SPECTRA OF SLIGHTLY DISTURBED LINEAR OSCILLATORS.*

By PHILIP HARTMAN.

The following theorem will be proved:

THEOREM. *Let* $q = q(t)$, *where* $0 \leqq t < \infty$, *be a real-valued, continuous function satisfying*

$$q(t) \rightarrow 0 \text{ as } t \rightarrow \infty. \tag{1}$$

Then the half-line $0 \leqq \lambda < \infty$ *is in the spectrum of the eigenvalue problem determined by*

$$x'' + (\lambda + q)x = 0, \tag{2}$$

by an arbitrary homogeneous boundary condition at $t = 0$, *say*

$$x(0) \cos \theta + x'(0) \sin \theta = 0, \qquad (0 \leqq \theta < \pi), \tag{3θ}$$

and by the (L^2)*-condition at* $t = \infty$,

$$\int^{\infty} x^2 dt < \infty. \tag{4}$$

That (2), (3θ) and (4) actually determine an eigenvalue problem is assured by (1); in fact, the boundedness of $q(t)$ from above is sufficient to this end; cf. [5], p. 252. It is known that the conclusion of the Theorem is valid if the assumption (1) is replaced by either the assumption

$$\int^{\infty} | q | dt < \infty, \tag{5_1}$$

[5], p. 264; or the assumptions

$$\int^{\infty} | dq | < \infty \text{ and } q \rightarrow 0 \text{ as } t \rightarrow \infty, \tag{5_2}$$

[8], p. 270, and Theorem (I), [9], p. 23; or the assumption

$$\int^{\infty} q^2 dt < \infty, \tag{5_3}$$

---

* Received June 3, 1948.

[3]. In [5], p. 264, the relation (1) as well as the condition (5_1) is assumed, but (1) is not needed in this situation. It is known that (5_1) implies that the half-line $0 \leqq \lambda < \infty$ is in the *continuous spectrum* and that no positive λ is in the point spectrum of a boundary value problem (2), (3θ) and (4). The same holds if (5_1) is replaced by (5_2), in virtue of the Theorem (I), [9], and the asymptotic formula [8], p. 270, for the solutions of (2) under the conditions (5_2). The same assertions hold if (5_1) is replaced by

$$(5_4) \qquad \int^{\infty} tq^2 dt < \infty,$$

[4], pp. 833-834.

It remains undecided whether or not every $\lambda \geqq 0$ is in the *continuous spectrum* under the assumption (1). On the other hand, it is known that (1) is compatible with a positive eigenvalue; cf. the examples constructed in [6], pp. 394-395 or [8], pp. 268-269.

The proof of the Theorem will be based on two different ideas; first, the constructions just referred to; second, the arguments used in [2] and [3], the latter being manifestations of the use of the Lebesgue-Toeplitz norm construction in [9], pp. 26-27, for locating points of the spectrum.

It will be convenient to deduce the theorem from several lemmas.

LEMMA 1. *Let $q = q(t)$, where $0 \leqq t < \infty$, be a continuous function satisfying* (1). *A given λ-value is in the spectrum of the eigenvalue problem* (2), (3θ) *and* (4), *for every θ, if there exists a continuous function $\phi = \phi(t)$, where $0 \leqq t < \infty$, with the property that the differential equation*

$$(6) \qquad y'' + (\lambda + \phi)y = 0$$

possesses a solution $y = y(t)$ satisfying

$$(7) \qquad y \to 0 \text{ and } y' \to 0 \qquad (t \to \infty)$$

and

$$(8) \qquad \int_0^{\infty} y^2 dt = \infty,$$

and, finally,

$$(9) \qquad \int_0^{\infty} y^2\phi^2 dt < \infty \text{ and } \int_0^{\infty} y^2 q^2 dt < \infty.$$

Proof. It can be assumed that there exists a value of $\theta = \theta^*$, where $0 \leqq \theta^* < \pi$, determining a boundary condition $(3\theta^*)$ corresponding to which the given λ-value is an eigenvalue. For otherwise it follows from the oscillation theorem of [2] that λ is in the spectrum belonging to every boundary condition (3θ), where $0 \leqq \theta < \pi$. Let $x = x^*(t)$ denote an eigenfunction belonging to λ and θ^*, so that $x = x^*(t) \not\equiv 0$ satisfies (2), $(3\theta^*)$ and (4). Then (1) and (4) imply

$$x^* \to 0 \text{ and } x^{*\prime} \to 0, \qquad (t \to \infty); \tag{10}$$

cf. [7], p. 13.

Suppose, if possible, that for some value θ, where $0 \leqq \theta < \pi$, the given λ is not in the corresponding spectrum. Then it is known ([5], p. 251) that the inhomogeneous equation

$$x'' + (\lambda + q)x = g \tag{11}$$

has a unique solution $x = X(t)$ satisfying (3θ) and (4) whenever $g = g(t)$ is a continuous function of class (L^2) on $0 \leqq t < \infty$,

$$\int_0^\infty g^2 dt < \infty. \tag{12}$$

Choose the function g to be $(q - \phi)y$, where ϕ and y are the functions satisfying (6)-(9). Then (12) follows from (9), while (11) becomes

$$x'' + (\lambda + q)x = (q - \phi)y. \tag{13}$$

Let $x = X = X(t)$ denote the solution of (13) satisfying (3θ) and (4). Then

$$X \to 0 \text{ and } X' \to 0 \qquad (t \to \infty); \tag{14}$$

cf. [7], p. 13 (although the proof given *loc. cit.* is for an homogeneous equation, it is also applicable to the inhomogeneous equation (11)) or cf. [1].

It is seen from (6) that $x = y(t)$ also is a solution of the inhomogeneous equation (13). Hence, the difference $x = X - y$ is a solution of the homogeneous equation (2). Furthermore, since X is, but y is not, of class (L^2), cf. (8), the difference $X - y$ is not. Accordingly, the solutions $x = x^*$ and $x = X - y$ are linearly independent. Consequently, (7), (10) and (14) lead to a contradiction, since the Wronskian of x^* and $X - y$ is a non-vanishing constant. Hence the proof of Lemma 1 is complete.

LEMMA 2. *Let $q = q(t)$, where $0 \leqq t < \infty$, be a continuous function satisfying* (1). *Then there exists a function $z = z(t)$, where $0 \leqq t < \infty$, possessing a continuous second derivative and having the properties that*

$$z > 0 \text{ and } z' < 0, \tag{15}$$

$$z \to 0 \text{ and } z' \to 0 \qquad (t \to \infty), \tag{16}$$

$$z'/z \to 0, \qquad (t \to \infty), \tag{17}$$

$$\int_0^\infty z^2 dt = \infty \text{ but } \int_0^\infty z^2 q^2 dt < \infty, \tag{18}$$

$$\int_0^\infty z'^2 dt < \infty \text{ and } \int_0^\infty z^2 (z'/z)'^2 dt < \infty, \tag{19}$$

and

$$\int_0^\infty (z'/z) \cos 2t dt \text{ is convergent.} \tag{20}$$

Proof. To simplify the construction, the function z defined below will be a "smooth" function, except for a sequence of "corners" tending to ∞. At a corner, the symbol z', z'' will represent either of the limits $z'(t \pm 0)$, $z''(t \pm 0)$, respectively. In view of the convexity of the function z^2, cf. (21) below, these corners can obviously be removed without influencing the above relations. (For the application to be made in Lemma 3, it is unnecessary to carry out this smoothing process; cf. the remarks following (29) below.)

Let $\alpha = \alpha_k$, where $k = 1, 2, \cdots$, denote a sequence of numbers such that

(i) $\alpha_1 > 1$,

(ii) $|q(t)| < 1/k$ if $t > \alpha_k$,

(iii) $\alpha_{k+1} > \alpha_k^2$

(so that the sequence α_k increases faster than a geometric progression),

(iv) α_k is an odd integral multiple of $\frac{1}{2}\pi$.

Let $\beta = \beta_k$, where $k = 1, 2, \cdots$, denote a sequence of numbers such that

(v) $\alpha_k < \beta_k < \alpha_{k+1}$,

(vi) $\beta_k \sim \alpha_k^2/k \qquad (k \to \infty)$,

(vii) β_k is an odd integral multiple of $\frac{1}{2}\pi$.

Since this lemma concerns large t-values, it will suffice to define the function $z(t)$ on $\beta_1 \leqq t < \infty$. In terms of the numbers α_k, β_k, let $z(t)$ be the positive continuous function determined by

$$\begin{aligned} &z^2(t) = t^{-2} \text{ if } \beta_k \leqq t \leqq \alpha_{k+1} \text{ and } k = 1, 2, \cdots \\ &z^2(t) \text{ is linear if } \alpha_k \leqq t \leqq \beta_k \text{ and } k = 2, 3, \cdots. \end{aligned} \tag{21}$$

For the sake of simplicity, α, β, A, B will be written in place of α_k, β_k, A_k, B_k, respectively. Then

$$z^2(\alpha) = \alpha^{-2} \text{ and } z^2(\beta) = \beta^{-2} \tag{22}$$

and

$$z^2(t) = At + B \qquad (\alpha \leqq t \leqq \beta), \tag{23}$$

where

$$A = A_k = -(\alpha^{-2} - \beta^{-2})/(\beta - \alpha) = -(\beta + \alpha)/\alpha^2\beta^2.$$

Hence, by (vi),

$$-A \sim k\alpha^{-4}, \qquad (k \to \infty). \tag{24}$$

The function $z(t)$ is positive, decreasing and satisfies the first relation in (16). Also,

$$2(z'/z) \text{ is either } A/(At+B) \text{ or } -2/t, \tag{25}$$

according as t $(\geqq \beta_1)$ is or is not in an interval of the type $\alpha \leqq t \leqq \beta$. In the first case, it is seen from (22), (23) and (24) that

$$A/(At+B) = O(k\alpha^{-4}\beta^2) = O(k^{-1}).$$

Consequently, (17) holds. The second limit relation in (16) is implied by (17) and the first part of (16).

There remains to verify the statements involving the integrals in (18), (19) and (20). It is seen from (22) and the second part of (21) that

$$\int_\alpha^\beta z^2 dt = \tfrac{1}{2}(\alpha^{-2} + \beta^{-2})(\beta - \alpha) \sim \tfrac{1}{2}\beta\alpha^{-2}, \qquad (k \to \infty). \tag{26}$$

Hence, the first relation in (18) follows from (vi), since

$$\int_{\beta_1}^\infty z^2 dt \geqq \sum_k \int_\alpha^\beta z^2 dt \geqq \text{const.} \sum_k 1/k = \infty.$$

On the other hand,

$$\int_{\beta_1}^\infty z^2 q^2 dt \leqq \int_{\beta_1}^\infty q^2 t^{-2} dt + \sum_k \int_\alpha^\beta q^2 z^2 dt.$$

The first integral on the right converges in virtue of (1). From (ii), (vi) and (26),

$$\int_\alpha^\beta q^2 z^2 dt < k^{-2} \int_\alpha^\beta z^2 dt = O(k^{-3}).$$

Thus, the second part of (18) is verified.

Similarly, it is seen from the first part of (21) that

$$\int_{\beta_1}^{\infty} z'^2 dt < \int_{\beta_1}^{\infty} t^{-2} dt + \sum_k \int_\alpha^\beta z'^2 dt,$$

and from (23) and (25) that

$$4\int_\alpha^\beta z'^2 dt = \int_\alpha^\beta A^2(At+B)^{-1} dt = A \log (At+B) \Big|_\alpha^\beta .$$

From (22) and (23), the last expression equals

$$A \log \alpha^2/\beta^2 \sim \text{const. } k\alpha^{-4} \log \alpha,$$

in view of (24) and (vi). Since

$$\sum_k k\alpha^{-4} \log \alpha < \infty,$$

the first relation in (19) follows. From (25),

$$(z'/z)' = O(z'/z)^2,$$

as $t \to \infty$. Hence, (17) shows that

$$z^2(z'/z)'^2 = z^2 O(z'/z)^2 = o(z'^2),$$

so that the second part of (19) is implied by its first part.

In view of (25),

$$\int_{\beta_1}^{\infty} (z'/z) \cos 2t\, dt = -\int_{\beta_1}^{\infty} t^{-1} \cos 2t\, dt + \sum_k \int_\alpha^\beta ((z'/z) + t^{-1}) \cos 2t\, dt,$$

provided that the series converges. Since

$$\int_\alpha^\beta t^{-1} \cos 2t\, dt = O(\alpha^{-1}),$$

it follows that

$$\sum_k \Big| \int_\alpha^\beta t^{-1} \cos 2t\, dt \Big| = O(\sum_k \alpha^{-1}) < \infty.$$

An integration by parts gives

$$2\int_\alpha^\beta (z'/z) \cos 2t\, dt = (z'/z) \sin 2t \Big|_\alpha^\beta - \int_\alpha^\beta (z'/z)' \sin 2t\, dt.$$

The first expression on the right vanishes, since α and β are integral multiples of $\frac{1}{2}\pi$. From (25),

$$-2\int_\alpha^\beta (z'/z)'\sin 2t\,dt = \int_\alpha^\beta A^2(At+B)^{-2}\sin 2t\,dt;$$

if the new variable $s = -(t+B/A)$ is introduced, the last integral becomes

$$-\int_{-\beta-B/A}^{-\alpha-B/A} s^{-2}\sin(2s+2B/A)\,ds.$$

From (22), (23) and (24),

$$-\alpha - B/A = -\alpha^{-2}/A \sim \alpha^2/k$$

and

$$-\beta - B/A = -\beta^{-2}/A \sim k,$$

by (vi). In view of the oscillatory character of $\sin(2s+2B/A)$ and the monotony of s^{-2},

$$\int_{-\beta-B/A}^{-\alpha-B/A} s^{-2}\sin(2s+2B/A)\,ds = O(\beta+B/A)^{-2} = O(k^{-2}).$$

Consequently,

$$\sum_k \left|\int_\alpha^\beta (z'/z)\cos 2t.dt\right| = O(\sum_k k^{-2}) < \infty.$$

This completes the proof of (20) and of Lemma 2.

Lemma 3. *If* $q = q(t)$, *where* $0 \leqq t < \infty$, *is a continuous function satisfying* (1) *and if* $\lambda > 0$, *then there exists a continuous function* $\phi(t)$ *satisfying the conditions of Lemma* 1.

Proof. It can be supposed that $\lambda = 1$, otherwise the unit of length on the t-axis is changed in the proportion $1 : \lambda^{\frac{1}{2}}$. Let $z = z(t)$ denote a function possessing a continuous second derivative and satisfying (15)-(20). In terms of $z(t)$, define functions $\chi = \chi(t)$ and $\phi = \phi(t)$ by

(27) $$\chi(t) = 2(z'/z)\cos t$$

and

(28) $$-\phi(t) = \chi^2\cos^2 t + \chi'\cos t - 3\chi\sin t,$$

respectively. Then ([8], p. 268)

(29) $$y = y(t) = \exp(\int_0^t \chi(s)\cos s\,ds)\cos t$$

is a solution of the differential equation (6), where $\lambda = 1$. It remains to show that (7), (8) and (9) hold.

It can be remarked that even if z is the function constructed in the proof of Lemma 2, so that z', z'' are discontinuous at $t = \alpha_k, \beta_k$, the functions (27) and (28) are still continuous, while (29) possesses a continuous second derivative. This is a consequence of the fact that α_k, β_k are odd integral multiples of $\frac{1}{2}\pi$.

In virtue of (27), the integral in the exp ( ) factor in (29) is

$$2\int_0^t (z'/z) \cos^2 s\, ds = \int_0^t (z'/z)\, ds + \int_0^t (z'/z) \cos 2s\, ds.$$

But (20) shows that the last integral tends to a finite limit as $t \to \infty$. Hence, the local absolute continuity of $\log z$ implies

$$\exp\left(\int_0^t \chi(s) \cos s\, ds\right) \sim C\, z(t), \qquad (t \to \infty), \tag{30}$$

where

$$C = z^{-1}(0) \exp\left(\int_0^\infty (z'/z) \cos 2t\, dt\right) > 0.$$

Consequently, the first part of (7) follows from (29), (30) and the first part of (16). A differentiation of (29) gives

$$y' = y\chi \cos t - \exp(\quad) \sin t,$$

so that the second part of (7) is implied by the first part of (7), (27) and (17), (30) and the first part of (16).

The monotony of z and the first part of (18) show that

$$\int_0^\infty z^2 \cos^2 t\, dt = \infty,$$

so that (8) follows from (29) and (30). Also, the second part of (9) follows from (30) and the second part of (18). Finally, in view of (28), the first part of (9) will be verified if it is shown that

$$\int^\infty \chi^4 z^2 dt < \infty, \int^\infty \chi'^2 z^2 dt < \infty \text{ and } \int^\infty \chi^2 z^2 dt < \infty. \tag{31}$$

Since (17) implies that $\chi \to 0$ as $t \to \infty$, the first of the relations (31) is a consequence of the last. On the other hand, the last part of (31) follows from (27) and the first part of (19). Finally, from (27),

$$\chi' = 2(z'/z)' \cos t - 2(z'/z) \sin t.$$

Hence, the second relation in (31) is implied by (19). This completes the proof of Lemma 3.

The Theorem now follows from Lemmas 1 and 3 and the fact that the spectrum is a closed set.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman, "The L^2-solutions of linear differential equations of second order," *Duke Mathematical Journal*, vol. 14 (1947), pp. 323-326.

[2] ——— and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[3] C. R. Putnam, "On the spectra of certain boundary value problems," *American Journal of Mathematics*, vol. 71 (1949), pp. 109-111.

[4] S. Wallach, "On the location of spectra of differential equations," *American Journal of Mathematics*, vol. 70 (1948), pp. 833-841.

[5] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1909), pp. 220-269.

[6] A. Wintner, "The adiabatic linear oscillator," *American Journal of Mathematics*, vol. 68 (1946), pp. 385-397.

[7] ———, "(L^2)-conections between the potential and kinetic energies of linear systems," *ibid.*, vol. 69 (1947), pp. 5-13.

[8] ———, "Asymptotic integrations of the adiabatic oscillator," *ibid.*, vol. 69 (1947), pp. 251-272.

[9] ———, "On the location of continuous spectra," *ibid.*, vol. 70 (1948), pp. 22-30.

BEST APPROXIMATE INTEGRATION FORMULAS; BEST APPROXIMATION FORMULAS.*

By ARTHUR SARD.[1]

Introduction. A criterion for determining best formulas of approximate integration is proposed in Section **1**. A criterion for determining best approximation formulas of a class of general type is proposed in Section **5**.

Particular best and nearly best integration formulas are given in Section **2**.

1. Best approximate integration formulas.

Consider approximations of an integral

$$I = \int_a^b x(t)\,dt$$

of the form

$$A = k_1 x(t_1) + k_2 x(t_2) + \cdots + k_p x(t_p),$$

where $k_1, \cdots, k_p;\ t_1, \cdots, t_p$ are constants independent of the function $x(t)$ and $a \leqq t_i \leqq b$, $i = 1, \cdots, p$. Suppose that A is such that the approximation is exact: $A = I$, whenever $x(t)$ is a polynomial in t of degree n. The non-negative integer n is fixed throughout our discussion.

Let $\mathcal{A}$ be a class of such approximations A. We shall suggest a criterion for determining which (if any) approximation in $\mathcal{A}$ is *best.* Our criterion will be relative to the integer n.

Put

$$R[x] = I - A;$$

$R[x]$ is a linear (that is, additive and continuous) functional on the space C_0 of functions $x(t)$ continuous on $a \leqq t \leqq b$ with norm

$$\| x \| = \max_{a \leqq t \leqq b} | x(t) | .$$

Furthermore $R[x]$ vanishes whenever $x(t)$ is a polynomial of degree n. It follows that there exists a function $k(t)$ such that

---

* Received January 16, 1948.

[1] The author gratefully acknowledges financial support received from the Office of Naval Research, Navy Department.

$$R[x]=\int_a^b x^{(n+1)}(t)k(t)\,dt, \tag{1.1}$$

whenever $x(t)$ is a function absolutely continuous n-th derivative. The function $k(t)$ may be taken as the following:

$$k(t')=R[\psi_{t'}]=-R[\phi_{t'}], \tag{1.2}$$

$$\psi_{t'}=\psi_{t'}(t)=\begin{cases}0 & \text{if } t\leqq t',\\ (t-t')^n/n! & \text{if } t>t',\end{cases} \tag{1.3}$$

$$\phi_{t'}=\phi_{t'}(t)=\begin{cases}(t-t')^n/n! & \text{if } t\leqq t',\\ 0 & \text{if } t>t'.\end{cases} \tag{1.4}$$

As a matter of fact $k(t')$ in (1.2) is an n-fold integral of a function of bounded variation.[2]

For each approximation in $\mathcal{A}$ there is a function k and a representation (1.1) of the remainder $R[x]$ corresponding to the approximation. We shall say that an approximation in $\mathcal{A}$ is *best* if the integral

$$J=\int_a^b k^2(t)\,dt$$

corresponding to it is a minimum in the set of all such integrals corresponding to the elements of $\mathcal{A}$.

The use of the word "best" may be justified as follows. The best element of $\mathcal{A}$ should be one for which $R[x]$ is smallest, in some sense. If J is a minimum, that part of (1.1) which is independent of $x(t)$ is small. Precisely put, Schwarz's inequality implies that

$$|R[x]|\leqq J^{1/2}\Big[\int_a^b x^{(n+1)}(t)^2\,dt\Big]^{1/2}, \tag{1.5}$$

providing that $x^{(n+1)}(t)^2$ is integrable. Thus if (1.5) is used to appraise $R[x]$, the best formula of $\mathcal{A}$, as defined above, gives the least appraisal.

Heretofore the appraisal of (1.1) by the theorem of the mean:

$$|R[x]|\leqq\int_a^b |k(t)|\,dt \sup_{a\leqq t\leqq b}|x^{(n+1)}(t)| \tag{1.6}$$

has often been used. In many situations the appraisal (1.5) seems preferable to (1.6); such situations are ones in which it is easier to estimate

---

[2] The results of this paragraph are due to F. Riesz, Peano, and Rémès. Cf. A. Sard, "Integral representations of remainders," *Duke Mathematical Journal*, vol. 15 (1948), pp. 333-345.

$$\int_a^b x^{(n+1)}(t)^2 dt$$

than sup $|x^{(n+1)}(t)|$.

If the appraisal (1.6) is used, the least appraisal will be obtained from that approximation in $\mathcal{A}$ (if any) for which

$$\int_a^b |k(t)|\, dt$$

is a minimum. Actually to determine such an approximation may be very complicated in practice; even in simple cases irrational values of the coefficients may be involved.

In the past a procedure sometimes followed has been to choose that approximation in $\mathcal{A}$ (if any) which is exact for a degree greater than n. Such an approximation permits an appraisal in terms of a higher derivative of $x(t)$ than the $(n+1)$-th providing that $x(t)$ possesses certain proper higher derivatives. However the function $x(t)$, or our knowledge of $x(t)$, may be such that an appraisal in terms of the higher derivative may not be possible or useful. For example the higher derivative may not exist; or our knowledge of $x(t)$ may be confined to a set of particular values of $x(t)$, contaminated by error, from which we can estimate or guess at properties of the $(n+1)$-th derivative but not a higher derivative. Alternatively the higher derivative may be so large that an appraisal in terms of it will be less powerful that an appraisal in terms of the lower derivative.

As a precise example, suppose that

$$(1.7) \qquad I = \int_0^6 x(t)\, dt, \qquad A = c_0 x(0) + c_1 x(1) + \cdots + c_6 x(6),$$

where $c_0, c_1, \cdots, c_6$ are constants; and suppose that $\mathcal{A}$ consists of all formulas A of the form (1.7) which are exact whenever $x(t)$ is a polynomial of degree 1. The Newton-Cotes formula is that formula of $\mathcal{A}$ which is exact whenever $x(t)$ is a polynomial of degree 6. The Newton-Cotes formula admits a representation (1.1) with $n = 6$, providing that $x^{(6)}(t)$ is absolutely continuous. But if a representation (1.1) with $n = 1$ is to be used, the Newton-Cotes formula is neither convenient nor apt. Calculators, aware of this fact, have followed alternative procedures; for example they have used that formula of $\mathcal{A}$ which results from an application of the trapezoidal rule to the integrals

$$\int_0^1, \int_1^2, \cdots, \int_5^6 x(t)\, dt.$$

The particular formula thus obtained leads to an appraisal (1. 5) which is more than twice the appraisal given by the best formula, in our sense.[3]

2. Particular best integration formulas. In this section $R[x]$ will stand for the remainder in the approximation of

$$\int_0^m x(t)\,dt$$

by a linear combination of the $m+1$ values of $x(t)$ at $t=0,1,\cdots,m$; the linear combination being such that the approximation is exact whenever $x(t)$ is a polynomial of degree n. Put $x_i = x(i)$. Then

$$R[x] = \int_0^m x(t)\,dt - (c_0x_0 + c_1x_1 + \cdots + c_mx_m) \tag{2.1}$$

$$= \int_0^m x^{(n+1)}(t)k(t)\,dt,$$

whenever $x^{(n)}(t)$ is absolutely continuous. Here m and n are integers, $m \geqq 1$, $n \geqq 0$; $c_0, c_1, \cdots, c_m$ are constants; k is defined by (1. 2).

For fixed m and n, let $\mathfrak{A}_{mn}$ be the class of all formulas (2. 1). We shall show in the next section that either there is a unique best formula in $\mathfrak{A}_{mn}$ or $\mathfrak{A}_{mn}$ is empty. The latter case occurs if and only if m is less than the largest even number contained in n.

Table 1 gives the coefficients $c_0, c_1, \cdots, c_m$ of the best formulas in all cases $n \leqq 3$, $m \leqq 6$ for which formulas exist. In each case Table 1 gives also the quantity

$$J = J_{mn} = \int_0^m k^2(t)\,dt \tag{2.2}$$

which enters in the appraisal (1. 5). In each of the best formulas

$$c_{m-i} = c_i \qquad i = 0, 1, \cdots, m; \tag{2.3}$$

---

[3] For the best formula, $c_0 = c_6 = 41/104$, $c_1 = c_5 = 118/104$, $c_2 = c_4 = 100/104$, $c_3 = 106/104$, and $J = 77/6240 = .0123$.

For the iterated trapezoidal rule, $c_0 = c_6 = 1/2$, $c_1 = c_2 = c_3 = c_4 = c_5 = 1$, and $J = 6/120 = .05$.

For the Newton-Cotes formula, $c_0 = c_6 = 41/140$, $c_1 = c_5 = 216/140$, $c_2 = c_4 = 27/140$, $c_3 = 272/140$, and $J = 27/350 = .077$.

The coefficients of the best formula and the corresponding value of J are given in Table 1, $m = 6$, $n = 1$. The comparative effectiveness of the best formula and the iterated trapezoidal rule is indicated by the square root of the ratio of the values of J, viz. $[6/120 \div 77/6240]^{1/2} = 2.01$. The numerical values of J can be obtained from equation (3. 6) with $p = 3$, $n = 1$.

accordingly certain coefficients c_i are not tabulated. The use of Table 1 is illustrated by the best formula cited in footnote 3.

A computer using (2.1) may determine an appropriate value of n as follows. He may estimate (by means of differences or in some other way) the products

$$J_{mn}\int_0^m x^{(n+1)^2}dt$$

for different n, and then choose the value of n corresponding to the least estimate.

TABLE 1.

Best Approximate Integration Formulas.

m	$c_0\Delta$	$c_1\Delta$	$c_2\Delta$	$c_3\Delta$	Δ	J_{mn}	
				$n=0$.			
1	1				2	1/12	= .083333
2	1	2			2	2/12	= .166667
3	1	2			2	3/12	= .25
4	1	2	2		2	4/12	= .333333
5	1	2	2		2	5/12	= .416667
6	1	2	2	2	2	6/12	= .5
				$n=1$.			
1	1				2	1/120	= .008333
2	3	10			8	1/160	= .00625
3	4	11			10	1/120	= .008333
4	11	32	26		28	1/105	= .009524
5	15	43	37		38	5/456	= .010965-
6	41	118	100	106	104	77/6240	= .012340
				$n=2$.			
2	1	4			3	1/1890	= .000529
3	3	9			8	11/8960	= .001228
4	21	76	46		60	11/12600	= .000873
5	112	379	289		312	73/69888	= .001045-
6	55	192	132	172	155	11/10850	= .001014
				$n=3$.			
2	1	4			3	1/9072	= .000110
3	3	9			8	13/17920	= .000725
4	2349	9932	4430		7248	6557/36529920	= .000179
5	29392	110209	76819		86568	61633/193912320	= .000318
6	1082811	4409946	2225043	4304484	3290014	210047/921203920	= .000228

TABLE 2.

NEARLY BEST FORMULAS.

m	$c_0\Delta$	$c_1\Delta$	$c_2\Delta$	$c_3\Delta$	Δ	J	Percentage by which J exceeds J_{mn}.
					$n = 1.$		
4	39	115	92		100	143/15000 $= .009533$	.10
5	40	112	98		100	11/1000 $= .011$	.32
6	40	112	98	100	100	31/2500 $= .0124$	.49
					$n = 2.$		
5	43	146	111		120	79/75600 $= .001045$	.04
6	355	1238	853	1108	1000	7097/7000000 $= .001014$	.003
					$n = 3.$		
4	97	412	182		300	10193/56700000 $= .000180$	.15
5	407	1529	1064		1200	1154043/3628800000 $= .000318$	.06
6	329	1341	675	1310	1000	31923/140000000 $= .000228$	.003
5	81	307	212		240	46987/145152000 $= .000324$	1.85-
6	33	134	67	132	100	1487/6300000 $= .000236$	3.52

In certain cases the coefficients c_i of the best formulas are somewhat cumbersome. For this reason the "nearly best" formulas of Table 2 may be of interest. The formulas of Table 2 are of the form (2.1); in each case the formula is exact for degree n and

$$J = \int_0^m k^2(t)\,dt$$

is close to the absolute minimum J_{mn}. Table 2 gives J and also the percentage by which J exceeds J_{mn}. Since the appraisal (1.5) involves the square root of J, the extent to which the appraisal is increased by the use of a nearly best formula instead of the best formula is indicated by one half the percentage increase of J over J_{mn}. In two cases, two alternative nearly best formulas are given.

The cited formulas transform to formulas of approximation of other integrals readily. Thus suppose that

$$t^* = \alpha t + \beta, \qquad \alpha > 0,$$

$$x^*(t^*) = x(t),$$

$$R^*[x^*] = \int_{a^*}^{b^*} x^*(t^*)\,dt^* - [c^*_0 x^*(t^*_0) + c^*_1 x^*(t^*_1) + \cdots + c^*_m x^*(t^*_m)]$$

$$= \int_{a^*}^{b^*} x^{*(n+1)}(t^*)\,k^*(t^*)\,dt^*,$$

$$J^* = \int_{a^*}^{b^*} k^{*2}(t^*)\,dt^*,$$

where $t^*_i = \alpha i + \beta$, $b^* = t^*_m$, $a^* = t^*_0$. Then $R^*[x^*] = \alpha R[x]$ if

$$c^*_i = \alpha c_i, \tag{2.4}$$

in which case $k^*(t^*) = \alpha^{n+1} k(t)$ and

$$J^* = \alpha^{2n+3} J. \tag{2.5}$$

Thus best and nearly best formulas for

$$\int_{a^*}^{b^*} x^*(t^*)\, dt^*$$

can be obtained from Tables 1 and 2 respectively merely by using (2.4).

The following inequality holds whenever $J_{m_1 n}$ and $J_{m_2 n}$ exist (that is, whenever the classes $\mathfrak{A}_{m_1 n}$ and $\mathfrak{A}_{m_2 n}$ are non-empty):

$$J_{m_1 n} + J_{m_2 n} \geqq J_{m_1 + m_2 n}. \tag{2.6}$$

For, a formula (2.1) with $m = m_1 + m_2$ can be constructed by combining (2.1) with $m = m_1$ and (2.1) with $m = m_2$ translated so as to refer to the integral from m_1 to $m_1 + m_2$. By (2.5) the translation leaves $J_{m_2 n}$ unchanged. Furthermore, J for the combined formula is the left side of (2.6). The inequality (2.6) follows, since $J_{m_1 + m_2 n}$ is the minimum of all J with $m = m_1 + m_2$.

The equality in (2.6) holds if and only if the coefficients c_i obtained when the formulas are combined are precisely the coefficients of the best formula for $m = m_1 + m_2$, since the minimum of J is taken on at a unique point.

3. Derivation of the formulas of section 2. We shall consider the case of an odd number of ordinates:

$$m = 2p.$$

The case of an even number of ordinates is treated similarly.

After a suitable translation of the t-axis, the remainder is of the form:

$$\begin{aligned} R[x] = \int_{-p}^{p} x(t)\, dt - [a_0 x_p + a_1 x_{p-1} + \cdots + a_{p-1} x_1 + c x_0 \\ + b_0 x_{-p} + b_1 x_{-p+1} + \cdots + b_{p-1} x_{-1}], \end{aligned} \tag{3.1}$$

where $R[x] = 0$ whenever $x(t)$ is a polynomial of degree n. The latter condition is satisfied if and only if $R[x] = 0$ for $x = 1$, $x = t$, $x = t^2, \cdots$, $x = t^n$; that is, if and only if

$$\begin{aligned}
&(3.2) \qquad \sum_{i=0}^{p-1}(p-i)^q(a_i+b_i)=2p^{q+1}/(q+1), \quad q \text{ even}, \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad 1 \leqq q \leqq n; \\
&(3.3) \qquad \sum_{i=0}^{p-1}(p-i)^q(a_i-b_i)=0, \qquad\qquad q \text{ odd}, \\
&(3.4) \qquad \sum_{i=0}^{p-1}(a_i+b_i)=2p-c.
\end{aligned}$$

Our problem is to determine (a_i, b_i, c) such that the $n+1$ constraints (3.2), (3.3), (3.4) are satisfied and such that the integral

$$J=\int_{-p}^{p}k^2(t)\,dt$$

is a minimum, where k is given by (1.2). In the process of solution, we shall prove that $a_i=b_i$, as one expects.

In order to calculate $k(t')$ we use (1.2) and (1.3) for $t' \geqq 0$; (1.2) and (1.4) for $t' \leqq 0$. By (1.2) and (1.3),

$$\begin{aligned}
k(t') &= (p-t')^{n+1}/(n+1)!-a_0(p-t')^n/n!, \qquad p-1 \leqq t' \leqq p, \\
k(t') &= (p-t')^{n+1}/(n+1)!-a_0(p-t')^n/n!-a_1(p-1-t')^n/n!, \\
&\cdot \\
&\cdot \qquad\qquad\qquad\qquad\qquad\qquad p-2 \leqq t' \leqq p-1, \\
&\cdot \\
k(t') &= (p-t')^{n+1}/(n+1)!-a_0(p-t')^n/n!-a_1(p-1-t')^n/n! \\
&\qquad -\cdots-a_{p-1}(1-t')^n/n!, \qquad 0 \leqq t' \leqq 1.
\end{aligned}$$

Put

$$H=H(a_0,\cdots,a_{p-1})=\int_0^p k^2(t')\,dt'.$$

Squaring each of the above expressions for k and collecting like terms, we see that

$$H=u-2v_ia_i+w_{ij}a_ia_j,$$

where a repeated index indicates a sum, $i,j=0,1,\cdots,p-1$, and

$$\begin{aligned}
u &= \int_0^p (p-t)^{2n+2}/(n+1)!^2\,dt = p^{2n+3}/(n+1)!^2(2n+3), \\
v_i &= \int_0^{p-i}(p-t)^{n+1}(p-i-t)^n/(n+1)!n!\,dt \\
&= \int_0^{p-i}(s+i)^{n+1}s^n/(n+1)!n!\,ds, \\
w_{ji}=w_{ij} &= \int_0^{p-j}(p-i-t)^n(p-j-t)^n/n!^2\,dt \\
&= \int_0^{p-j}(s+j-i)^ns^n/n!^2\,ds, \qquad\qquad j \geqq i.
\end{aligned}$$

A similar calculation, starting with (1.2), (1.4), shows that

$$\int_{-p}^{0} k^2(t')\,dt' = H(b_0, \cdots, b_{p-1}),$$

where H is precisely the same function as before. Hence

$$J = H(a_0, \cdots, a_{p-1}) + H(b_0, \cdots, b_{p-1}). \tag{3.5}$$

Since c does not enter in (3.5), the constraint (3.4) merely serves to determine c in terms of (a_i, b_i).

We wish to minimize the quadratic function (3.5) subject to the constraints (3.2), (3.3). If the constraints are satisfied by (a_i, b_i), they are also satisfied by (b_i, a_i). Furthermore J is unchanged if a_i and b_i are interchanged. Hence if J, subject to the constraints, assumes a minimum at a *unique* point (a_i, b_i), it must be true that $a_i = b_i$, $i = 0, 1, \cdots, p-1$. (Otherwise interchanging a_i and b_i would give another minimizing point.)

The constraints (3.2) (3.3) are either inconsistent or consistent. If inconsistent, the class of formulas $\mathfrak{A}_{mn}$ is empty. If the constraints are consistent, they may be used to eliminate from J certain of the variables (a_i, b_i). After this elimination J will be a quadratic function of the remaining variables, considered as independent; and J will be always positive. The positiveness of J may be seen as follows. If $J = 0$ for (a_i, b_i), then $k(t) = 0$ almost everywhere. Hence $R[x] = 0$, by (1.1), for all functions $x(t)$ with absolutely continuous n-th derivative. But we can construct such a function $x(t)$ which vanishes at integral points but whose integral is positive. For such an x, $R[x] \neq 0$.

We now prove that J takes on its minimum at a unique point (a_i, b_i). Suppose that q of the variables (a_i, b_i) are independent. If $q = 0$, (a_i, b_i) is determined uniquely. If $q > 0$, denote the independent variables by $y_1, \cdots, y_q$. Each y_r is one of the variables (a_i, b_i). Then $J = J(y) = u' - 2v'_r y_r + w'_{rs} y_r y_s$, say, where repeated indices are summed and $r, s = 1, \cdots, q$. It will be sufficient to prove that the quadratic form $w'_{rs} y_r y_s$ is positive definite, for then $J(y)$ will then have a unique critical point. Suppose the contrary. Then for some (y_r), say (y'_r), $w'_{rs} y'_r y'_s \leqq 0$ and $y'_r y'_r \neq 0$. Now $w'_{rs} y'_r y'_s < 0$ is impossible, else $J(\lambda y')$ would be negative for sufficiently large λ. Hence $w'_{rs} y'_r y'_s = 0$ and $J(\lambda y') = u' - 2v'_r y'_r \lambda$ for all real λ. Since J is never negative it follows that $v'_r y'_r = 0$ and $J(\lambda y') = u'$. But this is impossible. Let ρ be an integer such that $y'_\rho \neq 0$. Let $x(t)$ be a function which vanishes at integral values of t other than the integral value corresponding to y'_ρ as a coefficient in (3.1), and which takes on the value unity there. Then

$$R[x]=\int_{-p}^{p}x\,dt-y'_p\lambda$$

for all λ, by (3.1). This contradicts (1.5) since $J=u'$ is independent of λ.

Since $a_i=b_i$ at the minimum, our problem is equivalent to the following: To minimize

$$J=2H(a_0,\cdots,a_{p-1}) \tag{3.6}$$

subject to the constraints

$$\sum_{i=0}^{p-1}(p-i)^q a_i=p^{q+1}/(q+1),\qquad q\text{ even},\qquad 1\leqq q\leqq n. \tag{3.7}$$

For particular p and n, this problem can be solved directly; one may use the constraints to eliminate certain of the variables and then minimize the quadratic function of the remaining variables. For example, if $n=2$, $p=2$, (3.6) and (3.7) become

$$J/2=32/63-16a_0/9-37a_1/120+8a_0^2/5+31a_0a_1/60+a_1^2/20,$$
$$4a_0+a_1=8/3.$$

Eliminate a_1:

$$J/2=13/315-7a_0/30+a_0^2/3.$$

For a minimum, $a_0=7/20$. Then $a_1=19/15$, $J=11/12600$. These results are given in Table 1, $n=2$, $m=4$.

In deriving particular best formulas, an elementary fact about quadratic functions is useful: Suppose that $f=\gamma+2\gamma_i z_i+\gamma_{ij}z_iz_j$, where repeated indices are summed, γ_{ij}, γ_i, γ are constants; and the z_i are independent variables. At a point z_i at which the differential df vanishes, $f=\gamma+\gamma_i z_i$.

Table 1, $n=0$, may be extended to all values of m quite easily: $c_0=c_m=1/2$, all other $c_i=1$. A proof of this fact may be given along the following lines. First one proves the assertion for the case m even, working with the explicit function (3.6). Then one derives therefrom the assertion for m odd by using the inequality (2.6). In this way one avoids formal complications due to the constraints.

4. Convergence. Consider a fixed n and a sequence $m=n$, $n+1$, $n+2,\cdots$ of best approximations of

$$\int_a^b x(t)\,dt$$

using respectively $m+1$ equally spaced ordinates $x(a)$, $x(a+(b-a)/m)$, $\cdots,x(b)$, each approximation being exact for degree n. Let $x(t)$ be a function with absolutely continuous n-th derivative for which $x^{(n+1)}(t)^2$ is integrable. Let $R_m[x]$ be the remainder in the approximation involving $m+1$ ordinates. Then

$$|R_m[x]| \leqq [\int_a^b x^{(n+1)}(t)^2 dt]^{1/2} J_m^{1/2},$$

where

$$J_m = [(b-a)/m]^{2n+3} J_{mn}, \tag{4.1}$$

by (1.5) and (2.5). An appraisal of the convergence of J_m to zero as $m \to \infty$ is the following:

$$0 < J_m \leqq [(b-a)/m]^{2n+3}(\{m/n\} J_{nn} + K_n), \tag{4.2}$$

where $\{u\}$ is the largest integer contained in u and

$$K_n = \max (J_{nn}, J_{n+1\,n}, \cdots, J_{2n-1\,n}) - J_{nn}.$$

The relation (4.2) follows from (4.1) and (2.6).

Thus $R_m[x]$ is of order at most $1/m^{n+1}$.

5. Best approximation formulas.

One may consider a far more general situation than that of the preceding paragraphs.

Suppose that $R[x]$ *is any functional which is linear* (additive and continuous) *on the space* C_q *of functions* $x = x(t)$ *with continuous* q-th *derivative on* $a \leqq t \leqq b$, the norm being

$$\| x \| = \max_{i=0,1,\ldots,q} \max_{a \leqq t \leqq b} | x^{(i)}(t) | .$$

Suppose that $R[x]$ *vanishes whenever* $x(t)$ *is a polynomial of degree* $n \geqq q$. Then (1.1), (1.2) hold whenever $x(t)$ is a function with absolutely continuous n-th derivative. Furthermore k in (1.2) is an $(n-q)$-fold integral of a function of bounded variation.

We define the *modulus* M of $R[x]$ by the relation:

$$M^2 = | K | \int_K k^2(t)\, dt,$$

where K is the set in $a \leqq t \leqq b$ on which $k(t) \neq 0$ and $| K |$ is the measure of K.

Suppose that $\mathfrak{R}$ *is a class of functionals* $R[x]$ *of the above type.* We think of each $R[x]$ as a remainder. We say that R is a *best functional in the class* $\mathfrak{R}$ if the modulus of R is a minimum among the moduli of all the elements of $\mathfrak{R}$.

A justification of this definition is as follows. If $x^{(n+1)}(t)^2$ is integrable, Schwarz's inequality implies that

$$(5.1) \qquad |R[x]| \leqq M\, [(1/|K|) \int_K x^{(n+1)}(t)^2 dt]^{\frac{1}{2}},$$

since the integral in (1.1) may be taken over K instead of $[a, b]$. (The case $|K| = 0$ is ruled out as trivial, for then $R[x] = 0$.) The second member of (5.1) is the product of the modulus and the root-mean-square of $x^{(n+1)}(t)$ on K. As our choice of the best functional is to be independent of the particular function $x(t)$, we proceed as if the root-mean-squares of $x^{(n+1)}(t)$ on different sets K are comparable. The best functional, as just defined, gives the appraisal (5.1) with the least M.

The definition of best approximate integration formula given in Section 1 is consistent with the present definition. For in the cases considered in Section 1, $k(t)$ is piecewise a polynomial of degree $n + 1$. Accordingly K differs from $[a, b]$ by a finite number of points, $|K| = (b - a)$, and $M = (b - a)^{\frac{1}{2}} J^{\frac{1}{2}}$. Hence minimizing J is equivalent to minimizing M.

As a particular example, suppose that we wish to approximate the derivative $x'(1/4)$ in terms of a linear combination of certain of the values

$$(5.2) \qquad \cdots, x(-2), x(-1), x(0), x(1), x(2), \cdots;$$

that the approximation is to be exact for degree n_0, and that we will appraise the remainder by (5.1) with $n = n_0$. The more of the values (5.2) that we use, the smaller becomes the integral in M^2. At the same time, however, the larger $|K|$ may become. If $n_0 = 1$, the approximation using the fewest and closest values (5.2) seems to be best. If $n_0 = 2$, the approximation using the fewest and closest values (5.2) is not the best. Details are omitted here.

The functionals $R[x]$ considered in this section may be the remainders in many approximating processes, including the following: Approximate integration other than that considered in Section 1, Approximate integration of Stieltjes integrals, Interpolation, Extrapolation, Approximate differentiation, Approximation summation, Approximations by polynomials according to least squares.

Finally, suppose that $\mathfrak{R}$ is a class of linear *operations* $R[x]$ on C_q to a function space, say, C_0, the space of functions $y = y(u)$ continuous on $\bar{a} \leqq u \leqq \bar{b}$. Suppose that each $R[x]$ vanishes whenever $x(t)$ is a polynomial of degree $n \geqq q$. For each fixed u, then, $R[x]$ has a modulus M. We may define a best operation (if any) in $\mathfrak{R}$ as one which minimizes the integral of M^2 or, alternatively, the maximum of M^2, over $\bar{a} \leqq u \leqq \bar{b}$.

QUEENS COLLEGE.

CHARACTERIZATION OF A FIELD BY A SINGLE OPERATION.*

By S. Borofsky.

1. Introduction. A field is a system consisting of a set of elements and two binary operations with properties roughly equivalent to the usual properties of addition and multiplication in ordinary algebra.[1] These operations are distinct, that is, the sum of elements a and b cannot always equal their product. They are not completely independent, however, since they are connected by the distributive law $a(b+c)=ab+ac$. It is not surprising, therefore, that a single operation can be found, consisting of some combination of addition and multiplication, in terms of which the two field operations can be expressed.

N. Wiener[2] has shown that a field can be characterized by suitably ascribing properties to a single operation in such a way that this operation becomes the equivalent of $1-a/b$ in the field. If $b=0$, however, the result of this operation does not exist, so that the field fails to be closed under the operation. R. J. Levit[3] showed that a field could also be characterized by another operation, expressible in the form $a(1-b)$, under which the field is closed. To avoid some complexity in the definition and properties of addition and subtraction, Levit demonstrates the sufficiency of his postulates by deducing Wiener's from his own. However, merely by changing the primitive operation to $a(b-1)$, using a different set of postulates, it is an extremely simple matter to define the field operations and deduce immediately the necessary properties. This is done in paragraph 2.

In paragraph 3 we discuss other ways of defining the field operations. In paragraph 4 we show how certain single operations characterize subfields of a given field.

2. Characterization of a field. Let S be a system consisting of a set of elements, at least two in number, and an operation $a \circ b$. We assume:

A1. For all a, b in S $a \circ b$ exists and is a unique element of S.

---

* Received August 8, 1947.

[1] See, for instance, *A Survey of Modern Algebra* by Birkhoff and MacLane. Note that a field must contain at least two elements.

[2] *Transactions of the American Mathematical Society*, vol. 21 (1920), pp. 237-246.

[3] *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 426-40.

A2. $(a \circ b) \circ c = (a \circ c) \circ b$.

A3. If $x \circ a = x \circ b$ for all x in S, then $a = b$.

A4. There exists an element 0 such that $(a \circ 0) \circ (0 \circ b) = a$ for all a, b in S.

A5. There exists an element x_0 such that corresponding to any a, b in S with $a \neq x_0$ there is an x for which $a \circ x = b$.

For simplicity in form, the statement of the final postulate is deferred until we have deduced some further properties of the operation.

THEOREM 1. *If* $a \circ b = a \circ c$, $a \neq x_0$, *then* $b = c$.

Let x be any element of S. Then there is an element y such that $a \circ y = x$ (A5). Then

$$\begin{aligned} x \circ b &= (a \circ y) \circ b \\ &= (a \circ b) \circ y \quad \text{(A2)} \\ &= (a \circ c) \circ y \quad \text{(Hyp.)} \\ &= (a \circ y) \circ c \quad \text{(A2)} \\ &= x \circ c. \end{aligned}$$

Hence $b = c$ (A3).

COROLLARY. *The* x *in* A5 *is unique.*

THEOREM 2. *There is a unique* 0 *and a unique* x_0 *and they are equal.*

Choose any 0 and any x_0. Then there must be an element a such that $a \circ 0 \neq x_0$. For otherwise for every a we have $a = (a \circ 0) \circ (0 \circ x_0)$ by A4, so that every element equals $x_0 \circ (0 \circ x_0)$, contrary to the existence of at least two elements.

Let a be so chosen that $a \circ 0 \neq x_0$. Then, for any elements b, c we have $(a \circ 0) \circ (0 \circ b) = (a \circ 0) \circ (0 \circ c)$ by A4. Therefore, $0 \circ b = 0 \circ c$ (Theorem 1) for all b, c. If $0 \neq x_0$ this requires $b = c$ (Theorem 1), which is impossible.

Thus every 0 equals every x_0, which establishes the desired result.

THEOREM 3. $0 \circ a = 0$ for every a.

We have already seen (preceding proof) that $0 \circ b = 0 \circ c$ for all b, c. Letting $c = 0$ we have $0 \circ b = 0 \circ 0$. Hence $(0 \circ b) \circ 0 = (0 \circ 0) \circ 0$. But $(0 \circ b) \circ 0 = (0 \circ 0) \circ b$ by A2. Therefore $(0 \circ 0) \circ b = (0 \circ 0) \circ 0$.

If $0 \circ 0 \neq 0$, this requires $b = 0$ (Theorem 1) for all b, which is impossible. Hence $0 \circ 0 = 0$, so that $0 \circ b = 0$.

Theorem 4. *There is a unique x such that $a \circ x = 0$ for all a.*

Let $b \neq 0$ be a fixed element. Let $b \circ x = 0$ (A5).

If a is any element, let $b \circ y = a$ (A5). Then

$$\begin{aligned} a \circ x &= (b \circ y) \circ x \\ &= (b \circ x) \circ y \quad \text{(A2)} \\ &= 0 \circ y \\ &= 0 \quad \text{(Theorem 3)} \end{aligned}$$

The uniqueness of such an x is immediate from A3.

Definition 1. *The unique x of Theorem 4 we denote by* 1.

Theorem 5. $1 \neq 0$.

If $1 = 0$, then for every a, b

$$\begin{aligned} a &= (a \circ 1) \circ (0 \circ b) \quad \text{(A4)} \\ &= 0 \circ (0 \circ 0) \quad \text{(Theorem 4, Theorem 3)} \\ &= 0 \quad \text{(Theorem 3)} \end{aligned}$$

which is impossible.

Theorem 6.[4] *For every a, b with $a \neq 1$ there exists a unique x such that $x \circ a = b$.*

Let $c \neq 0$. If $c \circ a = 0$ then $c \circ a = c \circ 1$ (Theorem 4) so that $a = 1$ (Theorem 1). Hence $c \circ a \neq 0$.

Let $(c \circ a) \circ y = b$ (A5). Let $x = c \circ y$. Then
$$\begin{aligned} x \circ a &= (c \circ y) \circ a \\ &= (c \circ a) \circ y \quad \text{(A2)} \\ &= b. \end{aligned}$$

Suppose we also have $x' \circ a = b$. Let $c \circ z = x'$ (A5). Then

$$b = x' \circ a = (c \circ z) \circ a = (c \circ a) \circ z \quad \text{(A2)}.$$

But $b = (c \circ a) \circ y$. Hence $(c \circ a) \circ y = (c \circ a) \circ z$, so that $y = z$ (Theorem 1). Therefore $x' = c \circ y = x$.

Definition 2. $-a = a \circ 0$.

Theorem 7. $-0 = 0$.

Theorem 8. $-(-a) = a$.

---

[4] Not needed in this paragraph; used in paragraph 3.

For,
$$\begin{aligned} -(-a) &= (-a)\circ 0 \quad \text{(Definition 2)}\\ &= (a\circ 0)\circ 0 \quad \text{(Definition 2)}\\ &= (a\circ 0)\circ (0\circ b) \quad \text{(Theorem 3)}\\ &= a \quad \text{(A4)}. \end{aligned}$$

COROLLARY. $(-a)\circ 0 = a$.

THEOREM 9. $(-a)\circ b = -(a\circ b)$.

For,
$$\begin{aligned} -(a\circ b) &= (a\circ b)\circ \quad \text{(Definition 2)}\\ &= (a\circ 0)\circ b \quad \text{(A2)}\\ &= (-a)\circ b \quad \text{(Definition 2)}. \end{aligned}$$

DEFINITION 3. $ab = (-a)\circ((-1)\circ b)$.

THEOREM 10. $a1 = a$.

For,
$$\begin{aligned} a1 &= (-a)\circ((-1)\circ 1) \quad \text{(Definition 3)}\\ &= (-a)\circ 0 \quad \text{(Theorem 4)}\\ &= a \quad \text{(Corollary to Theorem 8)}. \end{aligned}$$

COROLLARY. $11 = 1$.

THEOREM 11. $(-a)b = -ab$.

For,
$$\begin{aligned} (-a)b &= (-(-a))\circ((-1)\circ b) \quad \text{(Definition 3)}\\ &= -[(-a)\circ((-1)\circ b)] \quad \text{(Theorem 9)}\\ &= -ab \quad \text{(Definition 3)}. \end{aligned}$$

THEOREM 12. $a0 = 0a = 0$.

For,
$$\begin{aligned} a0 &= (-a)\circ((-1)\circ 0) \quad \text{(Definition 3)}\\ &= (-a)\circ 1 \quad \text{(Corollary to Theorem 8)}\\ &= 0 \quad \text{(Theorem 4)}, \end{aligned}$$

and
$$\begin{aligned} 0a &= (-0)\circ((-1)\circ a) \quad \text{(Definition 3)}\\ &= -[0\circ((-1)\circ a)] \quad \text{(Theorem 9)}\\ &= -0 \quad \text{(Theorem 3)}\\ &= 0 \quad \text{(Theorem 7)}. \end{aligned}$$

THEOREM 13. *For any a, b with $a \neq 0$ there is a unique x such that $ax = b$.*

Since $a \neq 0$, therefore $-a \neq 0$ (Theorems 7, 8). Let $(-a)\circ y = b$ ($A5$).

Since $1 \neq 0$ (Theorem 5), therefore $-1 \neq 0$ and there exists an x such that $(-1)\circ x = y$ (A5). Then $ax = (-a)\circ((-1)\circ x)$ (Definition 3)
$$= (-a)\circ y = b.$$

If we also have $ax' = b$, then $ax = ax'$, so that

$$
\begin{aligned}
(-a) \circ ((-1) \circ x) &= (-a) \circ ((-1) \circ x') \quad \text{(Definition 3)} \\
(-1) \circ x &= (-1) \circ x' \quad \text{(Theorem 1)} \\
x &= x' \quad \text{(Theorem 1).}
\end{aligned}
$$

COROLLARY. *If* $ab = 0$ *then* $a = 0$ *or* $b = 0$.

For if $a \neq 0$ then $a0 = 0$ (Theorem 12) so that $a0 = ab$ and $0 = b$ (Theorem 1).

THEOREM 14. $(a \circ b)c = ac \circ b$.

For,
$$
\begin{aligned}
(a \circ b)c &= [-(a \circ b) \circ ((-1) \circ c)] \quad \text{(Definition 3)} \\
&= [(-a) \circ b] \circ [(-1) \circ c] \quad \text{(Theorem 9)} \\
&= [(-a) \circ ((-1) \circ c)] \circ b \quad \text{(A2)} \\
&= ac \circ b \quad \text{(Definition 3).}
\end{aligned}
$$

COROLLARY. $(ab)c = (ac)b$.

In Theorem 14 replace a by $-a$ and b by $(-1) \circ b$. Then

$$
\begin{aligned}
[(-a) \circ ((-1) \circ b)]c &= (-a)c \circ ((-1) \circ b) \\
(ab)c &= (-ac) \circ ((-1) \circ b) \quad \text{(Definition 3, Theorem 11)} \\
&= (ac)b \quad \text{(Definition 3).}
\end{aligned}
$$

DEFINITION 4. *If* $a \neq 0$, *then* a^{-1} *is the unique* x *satisfying* $ax = 1$.

DEFINITION 5.
$$
\begin{aligned}
a + b &= (-b) \circ (-ab^{-1}) \quad \textit{if } b \neq 0 \\
&= a \quad \textit{if } b = 0.
\end{aligned}
$$

THEOREM 15. $a + 0 = 0 + a = a$.

For $a = 0$ it is immediate. For $a \neq 0$

$$
\begin{aligned}
0 + a &= (-a) \circ (-0a^{-1}) \quad \text{(Definition 5)} \\
&= (-a) \circ 0 \quad \text{(Theorem 12)} \\
&= a \quad \text{(Corollary to Theorem 8).}
\end{aligned}
$$

THEOREM 16. $(-a) + a = a + (-a) = 0$.

For $a = 0$ it is immediate. For $a \neq 0$

$$
\begin{aligned}
(-a) + a &= (-a) \circ (-(-a)a^{-1}) \quad \text{(Definition 5)} \\
&= (-a) \circ (aa^{-1}) \quad \text{(Theorems 11, 8)} \\
&= (-a) \circ 1 \quad \text{(Definition 4).} \\
&= 0 \quad \text{(Theorem 4).}
\end{aligned}
$$

The second part of the theorem follows if we replace a by $-a$.

Theorem 17. $(b+c)a = ba + ca$.

If any of a, b, c is 0 the result is immediate. Suppose none is 0. Then

$$\begin{aligned}(b+c)a &= [(-c)\circ(-bc^{-1})]a \quad \text{(Definition 5)}\\ &= (-c)a\circ(-bc^{-1}) \quad \text{(Theorem 14)}\\ &= (-ca)\circ(-bc^{-1}) \quad \text{(Theorem 11)}\\ ba+ca &= (-ca)\circ(-(ba)(ca)^{-1}) \quad \text{(Definition 5;}\\ &\qquad \text{also by Corollary to Theorem 13 } ca \neq 0).\end{aligned}$$

Let $cx = b$ (Theorem 13). Then

$$\begin{aligned}(ba)(ca)^{-1} &= [(cx)a](ca)^{-1}\\ &= [(ca)x](ca)^{-1} \quad \text{(Corollary to Theorem 14)}\\ &= [(ca)(ca)^{-1}]x \quad \text{(Corollary to Theorem 14)}\\ &= 1x \quad \text{(Definition 4).}\end{aligned}$$

Also $$\begin{aligned}bc^{-1} = (cx)c^{-1} &= (cc^{-1})x \quad \text{(Corollary to Theorem 14)}\\ &= 1x \quad \text{(Definition 4).}\end{aligned}$$

Therefore, $(b+c)a = ba + ca$.

We now introduce the final postulate,

A6. $(1\circ a) + b = (1\circ b) + a$.

Theorem 18. $ab = ba$.

Taking $b = 1$ in A6, $$\begin{aligned}(1\circ a)+1 &= (1\circ 1)+a\\ &= 0+a \quad \text{(Theorem 4)}\\ &= a \quad \text{(Theorem 15).}\end{aligned}$$

But $$\begin{aligned}(1\circ a)+1 &= (-1)\circ[-(1\circ a)1^{-1}] \quad \text{(Definition 5)}\\ &= (-1)\circ[-(1\circ a)1] \quad \text{(Definition 4, Theorem 13,}\\ &\qquad \text{Corollary to Theorem 10)}\\ &= (-1)\circ[-(1\circ a)] \quad \text{(Theorem 10)}\\ &= 1a \quad \text{(Definition 3).}\end{aligned}$$

Thus $1a = a$.

Taking $a = 1$ in Corollary to Theorem 14, $(1b)c = (1c)b$, so that $bc = cb$.

Theorem 19. $a(bc) = (ab)c$.

For, $$\begin{aligned}a(bc) &= (bc)a \quad \text{(Theorem 18)}\\ &= (ba)c \quad \text{(Corollary to Theorem 14)}\\ &= (ab)c \quad \text{(Theorem 18).}\end{aligned}$$

THEOREM 20. $a+b=b+a$.

For $a=0$ in A6,
$$\begin{aligned}(1\circ 0)+b&=(1\circ b)+0\\ (-1)+b&=1\circ b \quad \text{(Definition 2, Theorem 15).}\end{aligned}$$

Also,
$$\begin{aligned}(-x)(-y)&=-[x(-y)] \quad \text{(Theorem 11)}\\ &=-[(-y)x] \quad \text{(Theorem 18)}\\ &=-(-yx) \quad \text{(Theorem 11)}\\ &=yx \quad \text{(Theorem 8)}\end{aligned}$$

so that $(-1)(-1)=11=1$ (Corollary to Theorem 10), from which $(-1)^{-1}=-1$.

We have, therefore,
$$\begin{aligned}b+(-1)&=[-(-1)]\circ[-b(-1)^{-1}] \quad \text{(Definition 5)}\\ &=[-(-1)]\circ[-b(-1)]\\ &=[-(-1)]\circ[-(-b)] \quad \text{(Theorems 18, 11, 10)}\\ &=1\circ b \quad \text{(Theorem 8)}\end{aligned}$$

Thus, $b+(-1)=(-1)+b$ for every b.

For $a\neq 0$ we have
$$\begin{aligned}a+b&=(-1)(-a)+(-ba^{-1})(-a)\\ &\qquad \text{(Theorems 18, 19; Definition 4; Theorem 10)}\\ &=[(-1)+(-ba^{-1})](-a) \quad \text{(Theorem 17)}\\ &=[(-ba^{-1})+(-1)](-a)\\ &=(-ba^{-1})(-a)+(-1)(-a) \quad \text{(Theorem 17)}\\ &=b+a\\ &\qquad \text{(Theorems 18, 19; Definition 4; Theorem 10).}\end{aligned}$$

For $a=0$ the result is immediate from Definition 5.

THEOREM 21. $(a+b)+c=a+(b+c)$.

We have
$$\begin{aligned}[a+(-1)]+b&=(1\circ a)+b \quad \text{(shown in Theorem 20)}\\ &=(1\circ b)+a \quad \text{(A6)}\\ &=[b+(-1)]+a \quad \text{(shown in Theorem 20).}\end{aligned}$$

For $c\neq 0$,
$$\begin{aligned}(a+c)+b&=[(-ac^{-1}+(-1))+(-bc^{-1})](-c)\\ &\qquad \text{(Theorems 17, 18; Definition 4; Theorem 10)}\\ &=[((-bc^{-1})+(-1))+(-ac^{-1})](-c) \quad \text{(as just shown)}\\ &=(b+c)+a\\ &\qquad \text{(Theorems 17, 18; Definition 4; Theorem 10).}\end{aligned}$$

For $c = 0$ the desired result is obvious.

Remark 1. The preceding theorems show that the system consisting of the elements of S and the operations of addition and multiplication, as defined, is a field. For the primitive operation we have

$$\begin{aligned} a \circ b &= 1a \circ b \quad \text{(Theorems 18,10)} \\ &= (1 \circ b)a \quad \text{(Theorem 14)} \\ &= a(1 \circ b) \quad \text{(Theorem 18)} \\ &= a[b + (-1)] \quad \text{(shown in Theorem 20)} \\ &= a(b - 1). \end{aligned}$$

Conversely, it is obvious that if we start with a field, with addition and multiplication as primitive operations, and define $a \circ b$ as $a(b - 1)$, then this operation satisfies A1-A6. Moreover, if we then use $a \circ b$ as a primitive operation and define addition and multiplication as in the preceding, then the two defined operations are exactly the same as the original operations.

Remark 2. The operation $a(1 - b)$ satisfies A1-A5; it also satisfies A6 when and only when the field has characteristic two. Thus, A6 is independent of A1-A5.

3. Other field operations. It is possible to define addition and multiplication in terms of the primitive operation $a \circ b$ differently from the way it was done in the preceding and still have the resulting system turn out to be a field. For convenience, let F_0 denote the field defined in **2** and let operations performed in this field be symbolized by $[\ \ ]_0$. Let α, β be any fixed elements, $\alpha \neq 0$. If we now define

$$\begin{aligned} a + b &= [a + b - \beta]_0 \\ ab &= [\alpha(a - \beta)(b - \beta) + \beta]_0 \end{aligned} \tag{1}$$

it is easily verified that the resulting system F is a field. In F the zero element for addition is β; the negative of a is $[-a + 2\beta]_0$, and $a - b = [a - b + \beta]_0$. Also, the unity element for multiplication in F is $[\beta + 1/\alpha]_0$ and the inverse of $a \neq \beta$ is $[\beta + 1/\{\alpha^2(a - \beta)\}]_0$. For $\beta = 0$, $\alpha = 1$, F is identical with F_0.

The definitions (1) express the operations of F in terms of those of F_0. Direct verification shows that we also have

$$\begin{aligned} [a + b]_0 &= a + b - \delta \\ [ab]_0 &= \gamma(a - \delta)(b - \delta) + \delta \end{aligned} \tag{2}$$

where $\delta = 0$, and $\gamma = [\beta + (1/\alpha^2)]_0$ is a non-zero element of F. Conversely, from (2) we can obtain (1) with $\alpha = \delta + (1/\gamma^2)$, if we assume F to be a field.

In terms of the operations of F_0, the primitive operation $a \circ b$ is expressible in the form $[a(b-1)]_0 = [ab - a]_0$. Transforming to the operations of F by means of (2), we have

$$a \circ b = \gamma(a-\delta)(b-\delta) + 2\delta - a. \tag{3}$$

It is natural to inquire whether (1) gives all possible ways of defining addition and multiplication in terms of the primitive operation $a \circ b$ so that the resulting system F is a field. A partial answer is given by the following:

THEOREM. *If addition and multiplication be defined in terms of* $a \circ b$ *in such a way that the resulting system* F *is a field and so that* $a \circ b$ *is expressible in the form*

$$a \circ b = (\lambda_1 ab + \mu_1 a + \nu_1 b + \sigma_1)/(\lambda_2 ab + \mu_2 a + \nu_2 b + \sigma_2), \tag{4}$$

where $\lambda_i, \mu_i, \nu_i, \sigma_i$ *are elements of* F, *then the operations of* F *are expressible in the form* (1).[5]

Suppose the conditions of the hypothesis satisfied. We show that $a \circ b$ must be expressible in the form (3) and that, consequently, $a + b$ and ab are given by (1).

To prove the second part first, suppose $a \circ b$ given by (3). Let β and i denote the zero and unity elements of F. We have

(a) $\gamma \neq \beta$. For if $\gamma = \beta$ then $a \circ b = 2\delta - a$, contrary to A3.

(b) $\delta \circ 1 = \delta$. Hence $\delta = 0$ (Theorem 4).

(c) $a \circ (\delta + i/\gamma) = 0$. Hence $1 = \delta + i/\gamma$ (Theorem 4).

(d) $[-a]_0 = a \circ 0$ (Definition 2).
$\quad = 2\delta - a$

(e) $[ab]_0 = [-a]_0 \circ ([-1]_0 \circ b)$ (Definition 3)
$\quad = (2\delta - a) \circ ((2\delta - 1) \circ b)$ (part d)
$\quad = (2\delta - a) \circ ((2\delta - \delta - i/\gamma) \circ b)$ (part c)
$\quad = \gamma(a-\delta)(b-\delta) + \delta$ by applying (3)

(f) for $a \neq 0$, $[a^{-1}]_0$ is the solution of $[ax]_0 = 1$, which is
$\delta + i/\{\gamma^2(a-\delta)\}$ by parts e, c

---

[5] Compare Levit, *loc. cit.*[3], Theorem I. The result above differs in nature from Levit's since here F is a particular field.

(g) for $b \neq 0$, $[a+b]_0 = [-b]_0 \circ [-ab^{-1}]_0$ (Definition 5)

$$
\begin{aligned}
&= (2\delta - b) \circ (2\delta - [ab^{-1}]_0) \quad \text{(part d)}\\
&= (2\delta - b) \circ (2\delta - (\gamma(a-\delta)([b^{-1}]_0 - \delta) + \delta)) \quad \text{(part e)}\\
&= (2\delta - b) \circ (2\delta - \gamma(a-\delta)(\delta + i/\{\gamma^2(b-\delta)\} - \delta) - \delta)\\
&\hphantom{=} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(part f)}\\
&= (2\delta - b) \circ (\delta - \{a-d\}/\{\gamma(b-\delta)\})\\
&= a + b - \delta \text{ from (3).}
\end{aligned}
$$

for $b = 0$ we have $[a+0]_0 = a = a + 0 - \delta$.

Thus, we have equations (2) and, consequently, (1).

We now show that if $a \circ b$ has the form (4) it must have the form (3).

Since $a \circ b$ exists and is unique for all a, b, the denominator in (4) can never equal β. This requires $\lambda_2 = \mu_2 = \nu_2 = \beta$. Hence $\sigma_2 \neq \beta$ and

$$a \circ b = \lambda ab + \mu a + \nu b + \sigma.$$

If $\lambda = \beta$ then $0 = 0 \circ b = \mu 0 + \nu b + \sigma$ (Theorem 3). This requires $\nu = \beta$. Thus $a \circ b = \mu a + \sigma$, contrary to A3. Therefore, $\lambda \neq \beta$.

Since $a \circ x = b$ is uniquely solvable for x whenever $a \neq 0$ (Corollary to Theorem 1), therefore $\lambda a + \nu$, the coefficient of x, can equal β only when $a = 0$. Therefore $\nu = -\lambda 0$.

Similarly, since $x \circ a = b$ is uniquely solvable for x whenever $a \neq 1$ (Theorem 6), we must have $\mu = -\lambda 1$.

From $0 \circ b = 0$ (Theorem 3) we now obtain $\sigma = 0 + 1\lambda 0$. Thus

$$a \circ b = \lambda ab - 1\lambda a - 0\lambda b + 0 + 1\lambda 0 = \lambda(a-0)(b-1) + 0.$$

From $(1 \circ 0) \circ 0 = 1$ (A4, Theorem 3) we have $\lambda^2(1-0)^3 + 0 = 1$. This gives $\lambda(1-0) = \pm i$.

If $\lambda(1-0) = i$, then $1 = 0 + i/\lambda$ and

$$a \circ b = \lambda(a-0)(b-0-i/\lambda) + 0 = \lambda(a-0)(b-0) + 0 + 0 - a$$

which is the desired form (3).

If $\lambda(1-0) = -i$, then $1 = 0 - i/\lambda$ and

$$a \circ b = \lambda(a-0)(b-0+i/\lambda) + 0 = \lambda(a-0)(b-0) + a. \tag{5}$$

In this case we now show that F has characteristic two so that this form is again the desired form.

From (5) we have, as before,

$$\begin{aligned}
1 &= 0 - i/\lambda \\
[-a]_0 &= a \\
[ab]_0 &= -\lambda(a-0)(b-0)+0 \\
[a^{-1}]_0 &= 0 + i/\{\lambda^2(a-0)\} \text{ for } a \neq 0 \\
[a+b]_0 &= b-a+0 \text{ for } b \neq 0 \\
&= a \text{ for } b = 0 \\
[(1 \circ a)+b]_0 &= a+b-1 \text{ for } b \neq 0.
\end{aligned}$$

Taking $a = 0$, $b \neq 0$ in A6, $[(1 \circ 0) + b]_0 = 1 \circ b$

$$\begin{aligned}
0+b-1 &= \lambda(1-0)(b-0)+1 \\
&= -i(b-0)+1 \\
&= -b+0+1.
\end{aligned}$$

Thus $b - 1 = -b + 1$ and F has characteristic two.

4. Characterization of subfields. If we have a group under the operation $a \cdot b$, a non-empty subset S is a subgroup if and only if it contains the solution of $x \cdot a = b$ whenever it contains a and b. Since a field can also be characterized by a single operation, we inquire whether a subfield can be characterized by closure under the inverse of this operation. With a mild restriction this is true.

Let F be a field and let $a \circ b$ be given by (3), where δ and γ are any fixed elements of F and $\gamma \neq 0$. Let $a \triangle b$ denote the solution of $x \circ b = a$, so that

$$a \triangle b = \delta + (a-\delta)/\{\gamma(b-\delta)-1\} \text{ for } b \neq \delta + 1/\gamma.$$

Similarly let $a \nabla b$ denote the solution of $b \circ x = a$, so that

$$a \nabla b = \delta + (a+b-2\delta)/\{\gamma(b-\delta)\} \text{ for } b \neq \delta.$$

For convenience, we say a subset S of F is closed under these operations if their results are in S whenever a and b are, provided b does not have the excluded values.

If a subfield S contains δ and $\delta + 1/\gamma$, then it contains $1/\gamma = (\delta + 1/\gamma) - \delta$, so that γ is also in S. Hence S is closed under the operation $a \triangle b$. Conversely, if a subfield S is closed under $a \triangle b$, then it must contain δ and $\delta + 1/\gamma$. For, if $\delta + 1/\gamma$ is neither 0 nor 1, then

$$1/(\gamma\delta + 1) = (0 \triangle 0) - (1 \triangle 0)$$

and

$$1/\{\gamma(1-\delta)-1\} = (1 \triangle 1) - (0 \triangle 1)$$

are also in S, from which it follows that δ and $\delta + 1/\gamma$ are in S. If $\delta + 1/\gamma$ is 0 or 1, then the same result follows from this fact and the fact that one of the elements $1/(\gamma\delta + 1)$, $1/\{\gamma(1 - \delta) - 1\}$ (whichever is appropriate) is in S.

A converse result is contained in the following:

THEOREM. *If subset S contains 0, 1, δ and $\delta + 1/\gamma$, and is closed under the operation $a \triangle b$, then S is a subfield.*

For a and b in S and $b \neq \delta + 1/\gamma$, we have

$$b \triangle b = \delta + \{b - \delta\}/\{\gamma(b - \delta) - 1\} \neq \delta + 1/\gamma$$
$$a \triangle (b \triangle b) = \gamma(a - \delta)(b - \delta) + 2\delta - a = a \circ b.$$

Since $a \circ (\delta + 1/\gamma) = \delta$, therefore S is closed under $a \circ b$.

Also $[(\delta + 1/\gamma) \circ (a \triangle \delta)] \triangle \delta = a + 1/\gamma$ is in S. Hence for any a, b in S with $a \neq \delta$, $b \triangle (a + 1/\gamma) = \delta + (b - \delta)/\{\gamma(a - \delta)\}$ is in S. By the preceding step, for $a \neq \delta$, $\delta + 1/\gamma + (b - \delta)/\{\gamma(a - \delta)\} = b \nabla a$ is in S.

It is now easily verified that the system consisting of S and the operation $a \circ b$ satisfies A1-A6. If we define $[ab]_s$ and $[a + b]_s$ by definitions 3 and 5, we have

$$[ab]_s = \gamma(a - \delta)(b - \delta) + \delta$$
$$[a + b]_s = a + b - \delta.$$

With these two operations, S is a field, by **2**.

It follows, as in **3**, that

$$a + b = [a + b - 0]_s$$
$$ab = [\alpha(a - 0)(b - 0) + 0]_s \text{ where } \alpha = \delta + 1/\gamma^2.$$

Since $[1 - 0]_s = 1 + \delta$, therefore $\epsilon = 1 + \delta$ is in S. Since $[\epsilon^{-1}]_s = \delta + 1/\gamma^2$, therefore α is in S.

Hence S is a subfield of F.

Remark. To establish the conclusion of the theorem it is necessary to assume that the subset S contains all four of the elements 0, 1, δ, $\delta + 1/\gamma$. That is, the fact that S is closed under $a \triangle b$ and contains three of these elements does not assure that it contains the fourth. This is shown by the following examples in each of which S is closed under $a \triangle b$ and contains the three elements indicated, but does not contain the fourth.

(a) F any field with at least three elements; δ different from 0 and 1; γ non-zero but otherwise arbitrary; S the set of elements of F different from δ. S contains $0, 1, \delta + 1/\gamma$.

(b) F any field of characteristic two with at least three elements; $\delta = 0$; γ different from 0 and 1; S the subset consisting of 0 and $1/\gamma$. S contains $0, \delta, \delta + 1/\gamma$.

(c) F the field of rational functions of x with coefficients in a given field of characteristic three; $\delta = 1$; $\gamma = 1/x$; S the subset consisting of $1, 1 + x, 1 - x$. S contains $1, \delta, \delta + 1/\gamma$.

(d) F the field of rational functions of the complex variable z; $\delta = 0$; $\gamma = z$; S the set of rational functions with a finite value for $z = 0$. S contains $0, 1, \delta$.

For the operation $a \nabla b$ we have a similar result.

THEOREM. *If subset S contains $0, 1, \delta$, and $\delta - 1/\gamma$ and is closed under the operation $a \nabla b$, then S is a subfield.*

Let a and b be in S and $b \neq \delta$. Then $\delta \nabla b = \delta + 1/\gamma$ is in S.

Also $a \nabla (\delta - 1/\gamma) = -a + 1/\gamma + 2\delta$. Replacing a in this by $a \nabla b$, we have $\delta + (\delta - a)/\{\gamma(b - \delta)\}$. Replacing b in the last expression by $-b + 1/\gamma + 2\delta$ for $b \neq \delta + 1/\gamma$, we have $\delta + (a - \delta)/\{\gamma(b - \delta) - 1\} = a \triangle b$.

By the preceding theorem, S is a subfield.

Remark. As above, examples can be given to show that S may be closed under $a \nabla b$ and contain three of the elements $0, 1, \delta, \delta - 1/\gamma$ without containing the fourth.

BROOKLYN COLLEGE.

ON A DECOMPOSITION INTO SINGULARITIES OF THETA-FUNCTIONS OF FRACTIONAL INDEX.*

By Aurel Wintner.

Introduction. The angular analogue of Cauchy's symmetric stable distributions leads to Fourier series which are formal generalizations of Jacobi's elliptic ϑ_3-function, the latter being the angular analogue of the limiting case of a symmetric Gaussian distribution (for details, cf. [2]). The Fourier series in question are those given by (3) and (2) below.

The object of this note is to determine the singularities of this function. This will be accomplished by developing the function into a series which puts the singularities into evidence. The series is of the Mittag-Leffler type but turns out to consist of branch points, instead of poles. The latter branch points are all logarithmic or all algebraic according as the "fractional" index is irrational or rational.

The final result is stated at the end of the paper.

1. If

$$0 < \lambda < 1 \tag{1}$$

and

$$0 < q < 1, \tag{2}$$

then the series

$$1 + 2 \sum_{n=1}^{\infty} q^{n^{\lambda}} \cos 2\pi n x \tag{3}$$

defines, for real x, an even function having the period 1 and possessing derivatives of arbitrarily high order. But the function cannot be regular-analytic at every real x, since (1) implies that the coefficients of (3) do not tend to 0 as fast as the terms of a convergent geometric progression. In what follows, the singularities of this function of x will be determined.

The corresponding question does not arise if (1) is replaced by $\lambda \geqq 1$. In fact, if $\lambda = 1$, then (3) becomes the elementary function

$$(1 - q^2)/(1 - 2q \cos 2\pi x + q^2)$$

(a Green function belonging to Laplace's equation), while if $\lambda > 1$, then the

---

* Received April 10, 1948.

coefficients of (3) tend to 0 faster than the terms of any geometric progression, and so (3) becomes a transcendental entire function of x.

In particular, if $\lambda = 2$, then (3) becomes Jacobi's elliptic ϑ_3 (a Green or "source" function belonging to Fourier's equation).

2. Suppose (1) and put, if $0 \leqq x < \infty$,

$$f(x) = \int_0^\infty e^{-t^\lambda} \cos xt \, dt \tag{4}$$

or (what, by Fourier's inversion, is the same thing)

$$\pi e^{-x^\lambda} = \int_0^\infty f(t) \cos xt \, dt. \tag{5}$$

Then, as shown in [1], the expansion

$$f(x) = \sum_{m=0}^\infty c_m / x^{m+1+\lambda}, \tag{6}$$

where $x^{m+1+\lambda} > 0$ and

$$(-1)^m m! c_m = \lambda \Gamma([m+1]\lambda) \sin(\tfrac{1}{2}\pi[m+1]\lambda), \tag{7}$$

holds for

$$0 < x < \infty. \tag{8}$$

The convergence of the series (6) at every point of the half-line (8) is part of the statement. Hence, if (6) is multiplied by $x^{1+\lambda} > 0$, it follows that $x^{1+\lambda} f(x)$, where $x > 0$, defines a transcendental entire function of $z = x^{-1}$.

It should be mentioned for later use that, as $x \rightarrow \infty$,

$$f(x) \sim c_0 / x^\alpha = O(x^{-\alpha}), \text{ where } \alpha = 1 + \lambda > 1. \tag{9}$$

This is seen from (6) and (7), or rather from the first approximation to the proof of (6).

3. For the present, let $f(x)$ be thought of as being defined by (4) on the line

$$-\infty < x < \infty, \tag{10}$$

rather than just on the half-line $0 < x < \infty$. Thus $f(x) = f(-x)$, and so, by (9),

$$f(x) = O(|x|^{-\alpha}), \text{ where } \alpha > 1. \tag{11}$$

It is trivial from (11), where $\pm x > 0$, and from (5), where $x > 0$, that the conditions for the applicability of Poisson's summation formula

$$\sum_{n=-\infty}^{\infty} F(x+n) = \sum_{n=-\infty}^{\infty} e^{2\pi inx} \int_{-\infty}^{\infty} F(t)\, e^{-2\pi int} dt$$

are satisfied by $F(x) = f(x)$ and also by $F(x) = f(px)$, where p is any positive constant. This means that the series

$$\sum_{n=-\infty}^{\infty} f(px + pn) \tag{12}$$

defines a function which has the period 1 and is represented by its Fourier series, which is

$$\sum_{n=-\infty}^{\infty} e^{2\pi inx} \int_{-\infty}^{\infty} f(pt)\, e^{-2\pi int} dt. \tag{13}$$

Since $f(t) = f(-t)$, the integral occurring (13) is identical with

$$2\int_{0}^{\infty} f(pt) \cos 2\pi nt\, dt = (2/p) \int_{0}^{\infty} f(t) \cos (2\pi nt/p)\, dt,$$

the constant p being positive. Hence, (5) shows that the series (13) can be contracted into

$$(2\pi/p)\,(1 + 2\sum_{n=1}^{\infty} q^{n^{\lambda}} \cos 2\pi nx),$$

if $q = q(p)$ is defined by $-\log q = (2\pi/p)^{\lambda}$.

4. What this proves is that the Fourier series of the function (12) results if the series (3) is multiplied by the constant $2\pi/p$, where

$$p = 2\pi/(-\log q)^{1/\lambda}. \tag{14}$$

Clearly, (14) defines a one-to-one mapping of the interval (2) on the half-line $0 < p < \infty$, which is the parameter range admitted in (12).

In order to make applicable the expansion (6), let (12), where $f(-t) = f(t)$, be written in the form

$$f(px) + \sum_{n=1}^{\infty} \{f(pn + px) + f(pn - px)\}, \tag{15}$$

and let x be restricted to the interval

$$0 < x < 1. \tag{16}$$

Since (15) or (3) has the period 1, the replacement of (10) by (16) involves no loss (except that $x = 0$ is now excluded).

Since (16) implies that both $n + x$ and $n - x$, hence both $pn + nx$ and $pn - nx$, are positive for $n = 1, 2, \cdots,$ and since (6) holds under the proviso (8), the expansion (6) can be inserted into (15) under the proviso (16). Hence, the function (15) can be written in the form

$$(17) \qquad f(px) + \sum_{n=1}^{\infty} (\sum_{m=0}^{\infty} \{ (c_m/(pn + px)^{m+\lambda+1} + c_m/(pn - px)^{m+\lambda+1} \}),$$

if x is on the interval (16). Finally, it is readily seen from (7), (1) and (6) that (17) can be rearranged into $2\pi/p$ times

$$(18) \qquad (2\pi)^{-1} \sum_{m=0}^{\infty} c_m p^{-m-\lambda} \{ \sum_{n=0}^{\infty} (n + x)^{-m-\lambda-1} + \sum_{n=1}^{\infty} (n - x)^{-m-\lambda-1} \}.$$

Accordingly, if λ, q are on their respective ranges (1), (2) and if $p = p(q, \lambda) > 0$ is defined by (14), then the function (3) of period 1 can, in terms of the constants (7), be represented in the form (18), where x and $1 - x$, hence every $n + x$ and $n - x$, are positive and all exponents refer to the real-valued (i. e., positive) determination of the respective multi-valued functions.

It is clear from the order of magnitude of the coefficients (7) that the expansion (18) of (3) is a "series in singularities," having the character announced in the Introduction.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. Wintner, "The singularities of Cauchy's distributions," *Duke Mathematical Journal*, vol. 8 (1941), pp. 678-681.

[2] ———, "On the shape of the angular case of Cauchy's distribution curves," *Annals of Mathematical Statistics*, vol. 18 (1947), pp. 589-593.

ON THE SPECTRA OF CERTAIN BOUNDARY VALUE PROBLEMS.*

By C. R. PUTNAM.

1. In the differential equation

$$y'' + (\lambda + q)y = 0, \tag{1}$$

let λ be a real parameter and let $q = q(x)$ be a real-valued, continuous function on the half-line $0 \leqq x < \infty$. The differential equation (1) is in the *Grenzpunktfall,* in the terminology of Weyl [4], p. 238, if for some λ (and hence for all λ) the differential equation (1) possesses at least one solution $y(x)$ not of class (L^2), that is,

$$\int_0^\infty y^2(x)\,dx = \infty.$$

In this case, the equation (1) and a boundary condition

$$y(0)\cos\alpha + y'(0)\sin\alpha = 0, \qquad 0 \leqq \alpha < \pi, \tag{2_α}$$

determine a boundary value problem for every fixed α. It is known [4], p. 251, that a value λ is not in the spectrum of the boundary value problem determined by (1) and (2_α) if and only if for *every* continuous function $f(x)$ on $0 \leqq x < \infty$ of class (L^2), the inhomogeneous differential equation

$$y'' + (\lambda + q)y = f$$

possesses a unique solution of class (L^2) satisfying the boundary condition (2_α). This characterization of the spectrum, together with an adaptation of the Lebesgue-Toeplitz construction, has been applied by Wintner [6], pp. 26-27, to the problem of locating points of the spectrum. It will be used below to prove the following

THEOREM. *Let $q = q(x)$ be a real-valued, continuous function on the half-line $0 \leqq x < \infty$ belonging to class (L^2), that is,*

$$\int_0^\infty q^2(x)\,dx < \infty. \tag{3}$$

Then the differential equation (1) *is in the Grenzpunktfall and the half-line*

---

* Received May 8, 1948.

$\lambda \geqq 0$ *belongs to the spectrum of the boundary value problem determined by the differential equation* (1) *and an arbitrary boundary condition* (2_a).

If assumption (3) is replaced by

$$\int_0^\infty xq^2(x)\,dx < \infty, \tag{3 bis}$$

then the stattement above is implied b ythe theorem in [3], pp. 833-834, which asserts that, in this case, the half-line $\lambda \geqq 0$ is in the *continuous* spectrum and contains no points of the point-spectrum. It remains undecided whether or not the corresponding statement concerning the continuous spectrum is valid under the milder restriction (3). That the corresponding statement concerning the point-spectrum is false is seen from the above Theorem and the example constructed in [5], pp. 394-395.

2. Proof of theorem. The theorem in [1], § 1, and the remarks following it *loc. cit.*, § 2, imply that (1) is in the *Grenzpunktfall* in virtue of the assumption (3).

It is therefore sufficient to prove that every positive value λ is a cluster point of the spectrum of the boundary value problem determined by (1) and some fixed boundary condition (2_a). For, according to [4], p. 251, the set of points of the spectrum consisting of the continuous spectrum and the set of cluster points of the point spectrum is independent of the boundary condition (2_a). Suppose, if possible, that a given $\lambda > 0$ is not a cluster point of the spectrum for a boundary value problem. Since (1) is in the *Grenzpunktfall,* there exist two distinct boundary conditions (2_{a1}) and (2_{a2}) such that λ is not an eigenvalue for either of the two boundary value problems determined by (1) and the boundary conditions (2_{a1}) and (2_{a2}), respectively.

Let $y = r_1(x)$ and $y = r_2(x)$ denote linearly independent solutions of the differential equation

$$y'' + \lambda y = 0 \tag{4}$$

satisfying the boundary conditions (2_{a1}) and (2_{a2}), respectively. In virtue of (3) and the fact that each function $r_k(x)$ is a linear combination of $\sin \lambda^{\frac{1}{2}}x$ and $\cos \lambda^{\frac{1}{2}}x$, it follows that the functions $q(x)r_k(x)$ are continuous and of class (L^2) for $k = 1, 2$. Clearly, the function $r_k(x)$ is a solution of the inhomogeneous differential equation

$$y'' + (\lambda + q)y = qr_k, \qquad (k = 1, 2), \tag{5_k}$$

satisfying the boundary condition (2_{ak}).

Since λ is not in the cluster spectrum and since λ is not an eigenvalue for either of the two boundary value problems determined by (1) and (2_{ak}), where $k=1,2$, it follows that λ is not in the spectrum for either of these boundary value problems. Hence, there exists a unique solution $y=y_k(x)$ of (5_k), of class (L^2), satisfying the boundary condition (2_{ak}). Define the functions $z_k(x)$ by

$$z_k(x)=y_k(x)-r_k(x), \qquad (k=1,2). \tag{6}$$

Then, for $k=1,2$, $z_k(x)$ is a solution of (1). The functions $z_k(x)$ are linearly independent. For suppose $c_1z_1+c_2z_2\equiv 0$ for some pair of constants c_1 and c_2, it follows from (6) that $c_1r_1+c_2r_2$ is of class (L^2). Consequently, $c_1=c_2=0$, since the function $y\equiv 0$ is the only solution of (4) of class (L^2), and r_1 and r_2 are linearly independent.

Since every solution $y=y(x)$ of (1) is a linear combination of z_1, z_2, so that

$$y=y(x)=c_1y_1(x)+c_2y_2(x)-(c_1r_1(x)+c_2r_2(x)),$$

it is seen that no non-trivial solution $y(x)$ of (1) is of class (L^2). This, however, contradicts the theorem proved in [2], p. 376, according to which every point λ not a cluster point of the spectrum is an eigenvalue for some boundary condition. This completes the proof of the theorem.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman, "On differential equations with non-oscillatory eigenfunctions," *Duke Mathematical Journal*, vol. 15 (1948), pp. 697-709.

[2] ——— and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[3] S. Wallach, "On the location of spectra of differential equations," *American Journal of Mathematics*, vol. 70 (1948), pp. 833-841.

[4] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 220-269.

[5] A. Wintner, "The adiabatic linear oscillator," *American Journal of Mathematics*, vol. 68 (1946), pp. 385-397.

[6] ———, "On the location of continuous spectra," *American Journal of Mathematics*, vol. 70 (1948), pp. 22-30.

TWIN CONVERGENCE REGIONS FOR CONTINUED FRACTIONS $b_0 + K(1/b_n)$, II.*

By W. J. THRON.

---

1. **Introduction.** In a recent paper [2],[1] which shall here be referred to as TCR, the author studied twin convergence regions for continued fractions of the form

$$b_0 + 1/b_1 + 1/b_2 + : \cdots. \tag{1.1}$$

Two regions [2] B_f and B_g are called *twin convergence regions for continued fractions* (1.1) if the two conditions

$$\text{(a)} \qquad \sum | b_n | = \infty,$$

$$\text{(b)} \qquad b_{2n} \varepsilon B_f, \qquad b_{2n+1} \varepsilon B_g, \qquad n \geqq 0,$$

insure convergence of fraction (1.1) which satisfies these conditions. Whenever at least one of the regions B_f or B_g is bounded away from the origin condition (a) is implied by condition (b).

Two regions B_f and B_g are called *best* twin convergence regions for continued fractions (1.1) if they are twin convergence regions and if moreover there do not exist twin convergence regions B'_f and B'_g such that $B'_f \supset B_f$, $B'_g \supset B_g$, where at least one of the relations $B'_f = B_f$, $B'_g = B_g$ fails to hold.

None of the convergence regions derived in TCR are best regions. In Section 2 of this paper a large class of best twin convergence regions is determined. These regions are improvements of the regions obtained in Theorem 6.1 of TCR.

In Section 3 two examples are given which show that an analogous improvement of Van Vleck's Criterion and of Theorem 7.3 of TCR is not possible.

We conclude this section by a remark on notation. Let B be a set and $f(z)$ a single-valued function defined for all $z \varepsilon B$, then $f(B)$ is understood

---

* Received December 1, 1947. Presented to the American Mathematical Society November 28, 1947. This paper contains the results of a research project undertaken by the author under a contract with the Office of Naval Research.

[1] The numbers in brackets refer to the bibliography at the end of this article.

[2] The term region is here used for open sets as well as for closures of such sets.

to be the set of all points $w = f(b)$ where $b \varepsilon B$. Thus, for example, $-1/B$ is the set of all numbers $-1/b$, $b \varepsilon B$.

2. A class of best twin convergence regions. Throughout this section we shall be concerned with pairs of regions B_f and B_g defined as follows:

$$
\begin{aligned}
&z = r \cdot e^{i\theta} \varepsilon B_f \text{ if } r \geqq f(\theta), \quad 0 < m_1 < f(\theta) < m_2,\\
&z = r \cdot e^{i\theta} \varepsilon B_g \text{ if } r \geqq g(\theta), \quad 0 < m_3 < g(\theta) < m_4,
\end{aligned} \tag{2.1}
$$

where $f(\theta)$ and $g(\theta)$ are of period 2π.

It was proved in TCR that:

A necessary condition for B_f and B_g to be twin convergence regions for continued fractions of the form (1.1) *is that*

$$f(\theta) g(\pi - \theta) \geqq 4, \textit{ for all } \theta.$$

One is thus led to the consideration of regions satisfying the condition

$$f(\theta) g(\pi - \theta) = 4. \tag{2.2}$$

It was further shown in TCR that:

If the regions B_f and B_g satisfy conditions (2.1) *and* (2.2) *and if in addition the complements of both regions are convex, then the values of all terminating continued fractions of the form*

$$b_0 + 1/b_1 + \cdots + 1/b_k$$

lie in the region $B_f/2$ provided $b_{2n} \varepsilon B_f$ and $b_{2n+1} \varepsilon B_g$ for all $n \geqq 0$.

It is thus of interest to find all functions $f(\theta)$ and $g(\theta)$ which satisfy the assumptions of this theorem. This is done in the following lemma.

LEMMA 2.1. *Two regions B_f and B_g satisfy conditions* (2.1) *and* (2.2) *and both have convex complements if and only if*

$$
\begin{aligned}
f(\theta) &= f_0 \cdot \exp\left(\int_{\frac{1}{2}\pi}^{\theta} \tan \alpha(\psi)\, d\psi \right),\\
g(\theta) &= 4/f_0 \cdot \exp\left(\int_{\frac{1}{2}\pi}^{\theta} \tan \alpha(\pi - \psi)\, d\psi \right),
\end{aligned}
$$

where $\alpha(\theta)$ is periodic with period 2π, is continuous and satisfies the two conditions

$$|\alpha(\theta)| < \tfrac{1}{2}\pi - \epsilon$$

$$|\alpha(\theta) - \alpha(\phi)| / |\theta - \phi| \leqq 1, \qquad \theta \neq \phi,$$

for all θ *and all* ϕ.

It is easily seen that, unless the curve defined by $r = f(\theta)$ has a unique tangent line at every point, at least one of the two regions under consideration does not have a convex complement. Thus $f'(\theta)$ exists for all θ. One also deduces from the convexity condition imposed on the regions that $|f'(\theta)/f(\theta)| \leqq m_2/m_1$. These results insure that

$$f(\theta) = f_0 \exp\left(\int_{\frac{1}{2}\pi}^{\theta} (f'(\psi)/f(\psi))\,d\psi\right)$$

if the integral is taken in the sense of Lebesgue [3, p. 368]. One now defines

$$\alpha(\theta) = \operatorname{Arc\,tan} f'(\theta)/f(\theta).$$

It is then clear that

$$|\alpha(\theta)| < \tfrac{1}{2}\pi - \epsilon, \text{ for all } \theta.$$

The function $f(\theta)$ can therefore be written as

$$f(\theta) = f_0 \exp\left(\int_{\frac{1}{2}\pi}^{\theta} \tan\alpha(\psi)\,d\psi\right).$$

Now let $\rho = t(\phi)$ be the equation of the tangent line to the curve $r = f(\theta)$ at the point $f(\theta)e^{i\theta}$. It is given by

$$t(\phi) = \cos\alpha(\theta) f(\theta)/\cos(\phi - \theta + \alpha(\theta)), \quad \theta - \alpha - \pi/2 < \phi < \theta - \alpha + \pi/2.$$

The region bounded by and interior to the curve $r = f(\theta)$ is therefore convex if and only if

$$t(\phi)/f(\phi) \geqq 1$$

for every θ and every ϕ in the corresponding range. This condition can be written as

$$\begin{aligned}
\frac{f(\theta)\cos\alpha}{f(\phi)\cos(\phi - \theta + \alpha)} &= \exp\left(\int_{\phi}^{\theta} \tan\alpha(\psi)\,d\psi - \log\cos(\phi - \theta + \alpha) + \log\cos\alpha\right. \\
&= \exp\left(\int_{\phi}^{\theta} [\tan\theta(\psi) - \tan(\psi - \theta + \alpha(\theta))]\,d\psi\right) \\
&= \exp\left(\int_{\phi}^{\theta} \frac{\sin(\alpha(\psi) - \psi + \theta - \alpha(\theta))}{\cos\alpha(\psi)\cdot\cos(\psi - \theta + \alpha(\theta))}\,d\psi\right) \geqq 1.
\end{aligned}$$

The last inequality holds if and only if for every θ

$$(\alpha(\psi) - \alpha(\theta) + \theta - \psi)/(\theta - \psi) \geqq 0$$

holds for almost all $\psi \neq \theta$, $\theta - \alpha - \pi/2 < \psi < \theta - \alpha \; ; \pi/2$. The above inequality is equivalent to

$$(\alpha(\theta) - \alpha(\psi))/(\theta - \psi) \leqq 1.$$

By going through the analogous argument for the function $g(\theta) = 4/f(\pi - \theta)$ one obtains the additional condition

$$(\alpha(\theta) - \alpha(\psi))/(\theta - \psi) \geqq -1.$$

The restrictions as to the range of ψ for these conditions can be omitted, since if these conditions are satisfied in "the small" they are certainly satisfied in "the large."

One now constructs a sequence $\{\theta_n\}$ which is everywhere dense on the real axis and which is such that

$$|\,\alpha(\theta_n) - \alpha(\theta_m)\,| \,/\, |\,\theta_n - \theta_m\,| \leqq 1 \text{ for all } n, m, n \neq m.$$

A new function $\beta(\theta)$ is then introduced by the definition

$$\beta(\theta_n) = \alpha(\theta_n),$$
$$\beta(\theta) = \lim_{\theta_{n_i} \to \theta} \beta(\theta_{n_i}).$$

This function is continuous and is equal to $\alpha(\theta)$ for almost all θ. But then

$$\log f(\theta) - \log f_0 = \int_{\frac{1}{2}\pi}^{\theta} \tan \beta(\psi)\, d\psi.$$

Finally since $\beta(\theta)$ is continuous it follows that

$$\begin{aligned}\beta(\theta) &= \operatorname{Arctan} d(\log f(\theta) - \log f_0)/d\theta = \operatorname{Arctan}\,(f'(\theta)/f(\theta)) \\ &= \alpha(\theta)\end{aligned}$$

for all θ. This completes the proof of the necessity. The sufficiency is verified without difficulty.

Since $\alpha(\theta)$ was found to be a continuous function all integrals appearing in the remainder of this section may be considered as Riemann integrals.

We now restrict ourselves to the consideration of pairs of regions B_f and B_g which in addition to conditions (2.1) and (2.2) satisfy

$$(2.3) \quad \begin{cases} f(\theta) = f_0 \exp\left(\int_{\frac{1}{2}\pi}^{\theta} \tan \alpha(\psi)\, d\psi \right), \text{ where } \alpha(\theta) \text{ is continuous} \\ \text{has period } 2\pi \text{ and is such that} \\ |\,\alpha(\theta)\,| < \tfrac{1}{2}\pi - \epsilon_1,\ |\,\alpha(\theta) - \alpha(\phi)\,| \,/\, |\,\theta - \phi\,| < 1 - \epsilon_2, \theta \neq \phi. \end{cases}$$

A pair of regions B_f and B_g which satisfies conditions (2.1), (2.2) and (2.3) will be called a *regular pair of regions.* The corresponding pair of functions $f(\theta)$ and $g(\theta) = 4/f(\pi - \theta)$ will be called a *regular pair of functions.*

LEMMA 2.2. *Let $f(\theta)$ and $g(\theta)$ be a regular pair of functions and let functions $k_f(b, z)$ and $k_g(b, z)$ be defined as follows*

$$w = k_f(b, z) = b \exp \left(\int_{\frac{1}{2}\pi}^{\arg b} (\tan z\alpha(\psi) - \tan \alpha(\psi))\, d\psi \right),$$

$$w = k_g(b, z) = b \exp \left(\int_{\frac{1}{2}\pi}^{\arg b} (\tan z\alpha(\pi - \psi) - \tan \alpha(\pi - \psi))\, d\psi \right),$$

then there exists a positive number μ and a region D, defined by $z \varepsilon D$ if

$$|\Re(z)| < 1/(1 - \eta), \text{ where } \eta = \tfrac{1}{2} \min (\epsilon_1, \epsilon_2),$$

and

$$|\Im(z)| < \mu,$$

such that:

(i) *For every b the functions k_f and k_g are holomorphic functions of z for all z in D.*

(ii) *The regions $k_f(B_f, z)$ and $k_g(B_g, z)$ form a regular pair of regions for every z in D.*

(iii) *For $\Re(z) = 0, z \varepsilon D$*

$$w \,\varepsilon\, k_f(B_f, z) \text{ if } |w| \geqq f_0,$$

$$w \,\varepsilon\, k_g(B_g, z) \text{ if } |w| \geqq 4/f_0.$$

(iv) *$z = 1 \varepsilon D$ and $k_f(b, 1) = b$, $k_g(b, 1) = b$.*

Property (i) is an immediate consequence of a known theorem in the theory of functions [1, p. 172].

Since the tangent of an imaginary number is imaginary one has for $\Re(z) = 0$

$$|k_f| = f_0 |b| / f_0 \text{ for } b \,\varepsilon\, B_f$$

and

$$k_g = 4 |b| / (f_0 g(\arg b)) \geqq 4/f_0 \text{ for } b \,\varepsilon\, B_g,$$

which proves statement (iii). Property (iv) is immediately verified by substitution.

It remains to show that (ii) holds. To that end we consider k_f as a function mapping the b-plane into the w-plane. The variable z plays the role of a parameter. The Jacobian of the mapping is

$$J = \partial(\,|\,w\,|\,, \arg w)/\partial(\,|\,b\,|\,, \arg b)$$

$$= (1 + \Im(\tan z\alpha(\arg b)))\exp\Big(\int_{\frac{1}{2}\pi}^{\arg b}[\Re(\tan z\alpha(\psi)) - \tan\alpha(\psi)]\,d\psi\Big).$$

Now

$$\Im(\tan(x + iy)) = \tanh y/(\cos^2 x + \sin^2 x \tanh^2 y).$$

Due to the restrictions imposed on $\alpha(\theta)\cos^2 x\alpha(\theta) > h > 0$ for all z satisfying $|\,\Re(z)| < 1/(1-\eta)$. One hence can find a $\mu_1 > 0$ such that for $z \,\varepsilon\, D(\mu_1)$, that is for $|\,\Im(z)| < \mu_1$ and $|\,\Re(z)| < 1/(1-\eta)$, and for all $b \neq 0$, $J \neq 0$. The transformation k_f has therefore a single-valued and differentiable inverse for all of these points. The same is true for the transformation k_g. Hence the boundaries of the regions $k_f(B_f, z)$ and $k_g(B_g, z)$ are the images of the boundaries of the regions B_f and B_g, respectively. For z real, $z \,\varepsilon\, D(\mu_1)$ the boundaries of the image regions are given by the equations

$$r = |\,k_f(f(\theta)e^{i\theta}, z)| \quad\text{and}\quad r = |\,k_g(g(\theta)^{i\theta}, z)\,|\,,$$

respectively. It is easily verified that these two functions form a regular pair. This remark completes the proof of (ii) for real z.

Now consider the mapping

$$w_2 = k_f(k_f^{-1}(w_1, z), z + \Delta z), \qquad w_1 \neq 0.$$

For every z in $D(\mu_1)$ there exists a δ such that for $|\,\Delta z\,| < \delta$ the mapping defined by this function differs by a given arbitrary small amount from the identity mapping with respect to displacement of points as well as with respect to change in directions. This is seen to be a consequence of the fact that for z in $D(\mu_1)$ the functions

$$\partial\,|\,k_f\,|/\partial\,|\,b\,|\,, \quad \partial\,|\,k_f\,|/\partial \arg b, \quad \partial \arg k_f/\partial\,|\,b\,|\,, \quad \partial \arg k_f/\partial \arg b,$$

are continuous functions of z and that the first and the fourth of these functions do not vanish for z in $D(\mu_1)$.

It follows that a $\mu > 0$, $\mu \leqq \mu_1$ can be found such that for all z satisfying the two conditions

$$|\,\Im(z)| < \mu, \qquad |\,\Re(z)| < 1/(1-\eta)$$

the regions $k_f(B_f, z)$ and $k_g(B_g, z)$ have both convex complements. One now verifies easily that these regions also satisfy conditions (2.1) and (2.2) and thus completes the proof of (ii) and the lemma.

Now let B_f and B_g be a regular pair of regions and let the elements of the continued fraction (1.1) satisfy the conditions

$$b_{2n} \varepsilon B_f, \qquad b_{2n+1} \varepsilon B_g, \qquad n \geqq 0.$$

Consider the continued fraction

$$k_f(b_0, z) + 1/k_g(b_1, z) + 1/k_f(b_2, z) + \cdots. \tag{2.4}$$

The approximants of this continued fraction are holomorphic functions of z for z in D. This is a consequence of Lemma 2.2, (i) and Lemma 2.2 of TCR.

For no value of z in D does any approximant of (2.4) take on a value on the open line segment from 0 to $if_0/2$. As was stated in the beginning of this section the values of the approximants of (2.4) lie in the region $k_f(B_f, z)/2$ for $z \varepsilon D$ (since $k_f(B_f, z)$ and $k_g(B_g, z)$ form a regular pair of regions) and for no z in D does the corresponding value region contain a point of the line segment in question.

It follows that for z in D the approximants of (2.4) form a normal family of holomorphic functions.

It was shown by Pringsheim that the continued fraction (1.1) converges if all its elements satisfy the condition $|b_n| \geqq 2$. By a simple transformation one finds that the two conditions

$$|b_{2n}| \geqq f_0, \qquad |b_{2n+1}| \geqq 4/f_0, \qquad n \geqq 0$$

insure convergence of the continued fraction (1.1). One now notes that for $\Re(z) = 0$, $z \varepsilon D$ the elements of the continued fraction (2.4) satisfy these conditions.

An application of the Stieltjes-Vitali Theorem then leads to the conclusion that the continued fraction (2.4) converges for all z in D. But $z = 1 \varepsilon D$ and hence the continued fraction (1.1) converges. This completes the proof of the following theorem.

THEOREM 2.1. *Let $\alpha(\theta)$ be a continuous function of period 2π which satisfies the two conditions*

$$|\alpha(\theta)| < \tfrac{1}{2}\pi - \epsilon_1,$$

$$|\alpha(\theta) - \alpha(\phi)| / |\theta - \phi| < 1 - \epsilon_2 \quad \theta \neq \phi, \quad \epsilon_1 > 0, \quad \epsilon_2 > 0.$$

Let

$$f(\theta) = f_0 \exp \int_{\frac{1}{2}\pi}^{\theta} \tan \alpha(\psi)\, d\psi, \qquad f_0 > 0.$$

Then the regions B_f *and* B_g *defined by:*

$$r \cdot e^{i\theta} \varepsilon B_f \text{ if } r \geqq f(\theta),$$
$$r \cdot e^{i\theta} \varepsilon B_g \text{ if } r \geqq 4/f(\pi - \theta)$$

are best twin convergence regions for continued fractions (1.1).

The following results are corollaries of this theorem.

COROLLARY 2.1. *Let* a *and* c *be real numbers satisfying the relation* $a > c \geqq 0$ *and let* γ *be an arbitrary real number. Then the continued fraction* (1.1) *converges if for all* $n \geqq 0$

$$| b_{2n} - ce^{i\gamma} | \geqq a$$

and

$$| b_{2n+1} - 4ce^{-i\gamma}/(a^2 - c^2) | \geqq 4a/(a^2 - c^2).$$

COROLLARY 2.2. *Let* c *be an arbitrary real number. Then the continued fraction* (1.1) *converges if for all* $n \geqq 0$

$$| b_n - c | \geqq (c^2 + 4)^{\frac{1}{2}}.$$

3. Two examples. Let A_n/B_n be the n-th approximant of the continued fraction (1.1) then

$$A_n/B_n = b_0 + \sum_{k=1}^{n} (-1)^{k-1}/(B_k B_{k-1}), \tag{3.1}$$

where $B_0 = 1$, $B_1 = b_1$ and for $n \geqq 2$

$$B_n = b_n B_{n-1} + B_{n-2}.$$

For $n \geqq 2$ one thus has

$$b_n = (B_n - B_{n-2})/B_{n-1}. \tag{3.2}$$

EXAMPLE 3.1. *The continued fraction of the form* (1.1) *with*

$$b_0 = 1, \qquad b_1 = e^{i\pi/4},$$

$$b_n = 2 \sin \tfrac{1}{2}\pi(1/(n+3) + 1/(n+2)) \exp(\tfrac{1}{2}i\pi(1 - 1/(n+3)(n+2))), \qquad n \geqq 2,$$

diverges.

Let

$$d_n = \pi \sum_{k=1}^{n} 1/(k+3).$$

Employing formula (3. 2) one easily verifies that the continued fraction with $B_n = e^{id_n}$, $n > 1$, is the one defined in the example. The divergence of this continued fraction follows immediately from relation (3. 1). One now notes that for the elements b_n of the example $\sum | b_n | = \infty$.

Van Vleck's convergence criterion states that the continued fraction (1. 1) converges if (a) $\sum | b_n | = \infty$ and if

$$\text{(b)} \qquad -\tfrac{1}{2}\pi + \epsilon < \arg b_n < \tfrac{1}{2}\pi - \epsilon, \qquad \epsilon > 0.$$

It follows from Example 3. 1 that the ϵ cannot be omitted in Van Vleck's criterion.

The following example shows that in Theorem 7. 3 of TCR the quantity δ, used there, cannot be set equal to zero.

EXAMPLE 3. 2. *The continued fraction of the form* (1. 1) *with* $b_0 = 1$,

$$b_{2n-1} = 2 + i(4 + 5s_{n-1}2^{n-1}),$$
$$b_{2n} = i(2^n s_n), \qquad n \geqq 1,$$

where $s_0 = 0$, $s_n = 2(s_{n-1} + 2^{1-n})$, $n \geqq 1$, *diverges.*

With the help of formula (3. 2) one shows that the continued fraction of the form (1. 1) whose B_n are defined by

$$B_{2n} = (i/2)^n,$$
$$B_{2n+1} = s_{n+1}(1 + 2i)(i)^n, \qquad n \geqq 0,$$

has as elements the numbers given in the example. Further

$$| B_{2n}B_{2n-1} | = 3^{1/2}s_n 2^{-n} < 4 \cdot 3^{-1/2}$$

since

$$0 < s_n 2^{-n} = s_{n-1}2^{1-n} + 4^{1-n} = \sum_{k=0}^{n-1} 4^{-k} < 4/3.$$

Hence the continued fraction diverges.

WASHINGTON UNIVERSITY.

BIBLIOGRAPHY.

1. L. Bieberbach, *Lehrbuch der Funktionentheorie*, 4th edition, vol. 1, Leipzig, 1934.
2. W. J. Thron, "Twin convergence regions for continued fractions $b_0 + K(1/b_n)$," *American Journal of Mathematics*, vol. 66 (1944), pp. 428-438.
3. E. C. Titchmarsh, *The Theory of Functions*, 2nd edition, Oxford, 1939.

STRUCTURE OF GENETIC ALGEBRAS.*

By R. D. Schafer.

I. M. H. Etherington has studied the non-associative algebras which arise in the symbolism of genetics (references [5] through [10]). In these he defines a class of algebras called *train algebras,* and proves in [7] that this class includes algebras called *special train algebras* which are defined by their structure rather than by any type of recurrence equation.

From the algebraic point of view the concept of train algebra appears to be too inclusive, in that an analysis of the structure of train algebras seems feasible only when the rank of the algebra is small. However, from the point of view of genetics the concept of special train algebra is certainly too narrow. For, although the *gametic* algebras for the fundamental types of symmetrical inheritance are special train algebras, the corresponding *zygotic* (*copular,* etc.) algebras are not necessarily special train algebras [7, p. 6, footnote].

We introduce a concept of *genetic algebra* which is intermediate between (commutative) train algebra and special train algebra. The definition is more satisfactory than that of special train algebra on two counts: the structure of the algebra is not postulated, and the *duplicate* of a genetic algebra is a genetic algebra. It follows from this latter fact that our genetic algebras include, not only the fundamental symmetrical gametic algebras, but also the zygotic (copular, etc.) algebras obtained from them by duplication. On the other hand, this new concept is restrictive enough for us to deduce a transparent structure theory for genetic algebras.

It is only fair perhaps to caution the reader that our interest in these algebras is entirely in the algebraic formalism, and that we can give no indication beyond Etherington's own remarks in [5] and [8] of their possible contribution to the study of genetics. Also we use the name "genetic algebra" with some misgivings. Our results are applicable to the algebras arising in genetics where inheritance is symmetrical in the sexes, and we abbreviate "genetic algebra of symmetrical inheritance" to "genetic algebra."

1. Preliminaries. The principal tool of our investigation of genetic algebras is the *transformation algebra* [1, § 2]. Let $\mathfrak{A}$ be a non-associative

---

* Received December 5, 1947.

algebra of order n over a field $\mathfrak{F}$. Then for a fixed element x in $\mathfrak{A}$ the correspondences

$$a \rightarrow ax = aR_x, \qquad a \rightarrow xa = aL_x, \qquad \text{for all } a \text{ in } \mathfrak{A},$$

are linear transformations on $\mathfrak{A}$ called the *right* and *left multiplications* R_x and L_x respectively. If $\mathfrak{M}$ is a subset of the total matric algebra $(\mathfrak{F})_n$ of all linear transformations on $\mathfrak{A}$, the *enveloping algebra* of $\mathfrak{M}$ is the algebra of all polynomials in the transformations in $\mathfrak{M}$ with coefficients in $\mathfrak{F}$. The enveloping algebra of the set which consists of the identity I in $(\mathfrak{F})_n$, together with the right and left multiplications of $\mathfrak{A}$, is the *transformation algebra* $T(\mathfrak{A})$ of $\mathfrak{A}$. Clearly any T in $T(\mathfrak{A})$ may be written in the form

$$(1) \qquad T = \alpha I + f(R_{x_1}, L_{x_1}, R_{x_2}, \cdots), \qquad \alpha \text{ in } \mathfrak{F},\ x_i \text{ in } \mathfrak{A}.$$

If $\mathfrak{B}$ is any linear subspace of $\mathfrak{A}$, the enveloping algebra of the set of right and left multiplications of $\mathfrak{A}$ which correspond to elements in $\mathfrak{B}$ is denoted by $\mathfrak{B}^*$. That is, T in $\mathfrak{B}^*$ has the form (1) with $\alpha = 0$, x_i in $\mathfrak{B}$. (It should be noted that the transformations in $\mathfrak{B}^*$ are linear transformations on $\mathfrak{A}$, although for compactness the notation does not indicate this.)

A *homomorphism* H of an algebra $\mathfrak{A}$ over $\mathfrak{F}$ into an algebra $\mathfrak{C}$ over $\mathfrak{F}$ is a linear mapping of $\mathfrak{A}$ into $\mathfrak{C}$ such that

$$(2) \qquad (ax)H = aH \circ xH \qquad \text{for all } a, x \text{ in } \mathfrak{A},$$

where $\circ$ denotes multiplication in $\mathfrak{C}$. The *kernel* of H is the set $\mathfrak{B}$ of all b in $\mathfrak{A}$ such that $bH = 0$; $\mathfrak{B}$ is an ideal of $\mathfrak{A}$. The homomorphism is *onto* $\mathfrak{C}$ in case, for any c in $\mathfrak{C}$, there exists some a in $\mathfrak{A}$ such that $c = aH$. The relationships between homomorphisms, ideals, and difference algebras are well-known. In case the homomorphism in question is from $\mathfrak{A}$ into the base field $\mathfrak{F}$, we use a functional notation:

$$(3) \qquad \omega\colon x \rightarrow \omega(x), \qquad x \text{ in } \mathfrak{A},\ \omega(x) \text{ in } \mathfrak{F},$$

and (2) becomes

$$(4) \qquad \omega(ax) = \omega(a)\omega(x), \qquad \text{for all } a, x \text{ in } \mathfrak{A}.$$

We call an algebra $\mathfrak{A}$ *nilpotent* in case there exists an integer t such that every product of t elements in $\mathfrak{A}$, no matter how associated, is zero. This is what Albert has recently called *strongly nilpotent* [2, p. 528; 3, p. 549]. He has defined a nilpotent algebra in the following way: every sequence $a_1, \cdots, a_k$ of k elements of $\mathfrak{A}$ defines a special product $a^{(k)}$ of order k by either of the formulas $a^{(i)} = a^{(i-1)}a_i$ or $a^{(i)} = a_i a^{(i-1)}$ for $i > 1$; if all special products of order k are zero and some special product of order $k - 1$ is not zero, Albert calls $\mathfrak{A}$ *nilpotent* of *index* k. Certainly a strongly nilpotent algebra is nilpotent by these definitions. However, an observation of Etherington [7, p. 2] shows the equivalence of the two notions: for if every special product of index k in $\mathfrak{A}$ is zero, then every product of $t = 2^{k-1}$ elements in $\mathfrak{A}$, no matter how associated, is zero. Thus the concept of a *strongly* nilpotent algebra is redundant.

A necessary and sufficient condition that an ideal $\mathfrak{B}$ of a non-associative algebra $\mathfrak{A}$ be nilpotent is that the associative algebra $\mathfrak{B}^*$ be nilpotent [2, Lemma 5].

If a non-associative algebra $\mathfrak{A}$ is homomorphic to a semi-simple algebra (direct sum of simple algebras), there is an ideal $\mathfrak{R}$ of $\mathfrak{A}$, called the *radical* of $\mathfrak{A}$, such that $\mathfrak{A} - \mathfrak{R}$ is semi-simple and $\mathfrak{R}$ is contained in every ideal $\mathfrak{B}$ of $\mathfrak{A}$ such that $\mathfrak{A} - \mathfrak{B}$ is semi-simple. It is an immediate consequence of [2, Theorem 6] that any nilpotent ideal of $\mathfrak{A}$ is contained in the radical of $\mathfrak{A}$.

2. Baric algebras. A non-associative algebra $\mathfrak{A}$ of order n over a field $\mathfrak{F}$ is called *baric* in case it has a non-trivial representation of degree one—that is, in case there is a homomorphism (3) of $\mathfrak{A}$ into $\mathfrak{F}$ such that for some x_0 in $\mathfrak{A}$ we have $\omega(x_0) \neq 0$. It follows that ω is a homomorphism of $\mathfrak{A}$ onto $\mathfrak{F}$ since, for any α in $\mathfrak{F}$, we have $\omega(\alpha x_0/\omega(x_0)) = \alpha$. We call $\omega(x)$ the *weight* of x, and ω the *weight function* of $\mathfrak{A}$.

We denote the kernel of the homomorphism ω by $\mathfrak{N}$. Then a necessary and sufficient condition that a non-associative algebra $\mathfrak{A}$ be a baric algebra is that $\mathfrak{A}$ contain an ideal $\mathfrak{N}$ such that

$$\mathfrak{A} - \mathfrak{N} \cong \mathfrak{F}. \tag{5}$$

Thus any non-associative algebra $\mathfrak{N}$ of order $n - 1$ over $\mathfrak{F}$ gives rise to a baric algebra $\mathfrak{A}$ of order n over $\mathfrak{F}$ if we adjoin an element u to $\mathfrak{N}$ in any fashion such that the elements $u^2 - u$, uz, and zu are in $\mathfrak{N}$ for all z in $\mathfrak{N}$ (a trivial construction).

In a *gametic algebra* $\mathfrak{A}$ we take a basis $u_1, u_2, \cdots, u_n$ denoting the gametic types involved in some genetical situation of symmetrical inheritance [5, § 6]. If γ_{ijk} is the probability that an arbitrary gamete produced by an individual of zygotic type $u_i u_j$ ($= u_j u_i$) be of gametic type u_k, we have

$$u_i u_j = \Sigma_k \gamma_{ijk} u_k \qquad (i, j = 1, \cdots, n) \tag{6}$$

subject to the conditions

$$(7) \qquad \Sigma_k \gamma_{ijk} = 1 \qquad (i, j = 1, \cdots, n).$$

Etherington points out that equations (6) and (7)—not assuming commutativity—imply that $\mathfrak{A}$ is a baric algebra with weight function

$$(8) \qquad \omega \colon x = \Sigma_i \xi_i u_i \to \omega(x) = \Sigma_i \xi_i, \qquad \xi_i \text{ in } \mathfrak{F}.$$

The converse is also true: given any baric algebra $\mathfrak{A}$ there is a basis $u_1, u_2, \cdots, u_n$ of $\mathfrak{A}$ whose multiplication table (6) is subject to the conditions (7). Also the defining homomorphism ω has the form (8). For let $v_2, \cdots, v_n$ be a basis of the ideal $\mathfrak{N}$ of $\mathfrak{A}$. There exists an element u in $\mathfrak{A}$, but not in $\mathfrak{N}$, of weight 1, so that $\mathfrak{A}$ has basis $u, v_2, \cdots, v_n$. Write $u_1 = u, u_i = u + v_i$ $(i = 2, \cdots, n)$. Then

$$(9) \qquad \omega(u_i) = 1 \qquad (i = 1, \cdots, n).$$

Moreover, $\mathfrak{A}$ has the basis $u_1, u_2, \cdots, u_n$ satisfying (6) for some γ_{ijk} in $\mathfrak{F}$. Now (4) and (9) imply that $1 = \omega(u_i)\omega(u_j) = \omega(u_i u_j) = \Sigma_k \gamma_{ijk}\omega(u_k) = \Sigma_k \gamma_{ijk}$ so that (7) holds. Also (8) follows from (9).

If $\mathfrak{A}$ is a baric algebra, then $T(\mathfrak{A})$ is also a baric algebra. For if $\mathfrak{A}$ has weight function ω, a weight function θ for $T(\mathfrak{A})$ is defined by

$$(10) \qquad \theta(T) = \omega(uT),$$

where u is any element of weight 1 in $\mathfrak{A}$. That θ is well-defined by (10) is clear since, if T in $T(\mathfrak{A})$ is written in the form (1), we have equivalently

$$(11) \qquad \theta(T) = \alpha + f(\omega(x_1), \omega(x_1), \omega(x_2), \cdots).$$

Then θ is linear by (10), a homomorphism by (11), and non-trivial since $\theta(I) = 1$.

3. Genetic algebras. Etherington has investigated the non-commutative aspects of some of the concepts he has introduced. However, since any algebra encountered in genetics may be taken to be commutative [8, p. 26], we shall assume the commutative law in all that follows.

In a commutative algebra $\mathfrak{A}$ we have $L_x = R_x$ for all x, so that we may write T in $T(\mathfrak{A})$ in the form

$$(12) \qquad T = \alpha I + f(R_{x_1}, R_{x_2}, \cdots), \qquad \alpha \text{ in } \mathfrak{F},\ x_i \text{ in } \mathfrak{A}.$$

The *characteristic function* $|\lambda I - T|$ of T in (12) has coefficients which are polynomials in α and the coordinates of the x_i, polynomials which depend both on the function f and on elements of $\mathfrak{F}$ which are independent of T (that is, scalars completely determined by $\mathfrak{A}$).

We call a commutative baric algebra $\mathfrak{A}$ over $\mathfrak{F}$ with weight function ω a *genetic algebra* in case the coefficients of the characteristic function of T in (12), insofar as they depend on the x_i, depend only on the weights $\omega(x_i)$. That is, these coefficients are polynomials in α and the $\omega(x_i)$ having coefficients which involve certain elements of $\mathfrak{F}$ determined by $\mathfrak{A}$ in combinations determined by f. (Note: the fact that for a given T in $T(\mathfrak{A})$ the expression (12) is not unique has no bearing on our definition.)

This definition is an extension of Etherington's definition of train algebra [5, § 4]. Define the *right powers* x^k of x in $\mathfrak{A}$ by $x^1 = x$ and

$$x^k = xR_x^{k-1} \qquad (k = 2, 3, \cdots). \tag{13}$$

(Since we assume $\mathfrak{A}$ commutative, right powers and similarly defined *left powers* are equal.) Then a commutative baric algebra $\mathfrak{A}$ over $\mathfrak{F}$ with weight function ω is called a *train algebra* in case the coefficients of the (right) rank equation [4, § 19], insofar as they depend on x, depend only on $\omega(x)$. That is, there exist elements $\beta_1, \cdots, \beta_{r-1}$ in $\mathfrak{F}$ such that

$$x^r + \beta_1\omega(x)x^{r-1} + \cdots + \beta_{r-1}[\omega(x)]^{r-1}x = 0 \tag{14}$$

for all x in $\mathfrak{A}$, where x^k is the right power (13).

THEOREM 1. *A genetic algebra $\mathfrak{A}$ over $\mathfrak{F}$ is a train algebra.*

Let $T = R_x$ in (12), and write $\omega(x) = \xi$. Then, since the coefficients of the characteristic function

$$\phi(\lambda) = | \lambda I - R_x |$$

of R_x are homogeneous polynomials in the coordinates of x, we have, by the definition of a genetic algebra,

$$\phi(\lambda) = \lambda^n + \gamma_1\xi\lambda^{n-1} + \cdots + \gamma_n\xi^n, \tag{15}$$

for some $\gamma_1, \cdots, \gamma_n$ in $\mathfrak{F}$. Now (15) factors in a finite extension $\mathfrak{K}$ of $\mathfrak{F}$ as

$$\phi(\lambda) = (\lambda - \lambda_1\xi)(\lambda - \lambda_2\xi) \cdots (\lambda - \lambda_n\xi), \qquad \lambda_i \text{ in } \mathfrak{K}. \tag{16}$$

If

$$\lambda^r + \psi_1\lambda^{r-1} + \cdots + \psi_{r-1}\lambda \tag{17}$$

is the rank function of $\mathfrak{A}$, ψ_j a homogeneous polynomial of degree j in the coordinates of x, then (17) divides $\lambda\phi(\lambda)$ [4, § 19]. The λ_i in (16) may then be ordered so that (17) equals $\lambda(\lambda-\lambda_1\xi)\cdots(\lambda-\lambda_{r-1}\xi)$, from which it follows that

$$\psi_j=(-1)^j\xi^j\Sigma\lambda_{i_1}\cdots\lambda_{i_j}, \qquad j=1,2,\cdots,r-1.$$

The rank equation is (14) with

$$\beta_j=(-1)^j\Sigma\lambda_{i_1}\cdots\lambda_{i_j}, \qquad j=1,2,\cdots,r-1,$$

in $\mathfrak{F}$; $\mathfrak{A}$ is a train algebra.

In a non-associative algebra in which powers of a single element are not necessarily associative, the concept of a nilpotent element may be defined variously. Here we shall call an element z *nilpotent* in case there exists an integer k for which the right power $z^k=0$. In a train algebra $\mathfrak{A}$ the kernel $\mathfrak{N}$ of the weight function ω then has an easy characterization: $\mathfrak{N}$ consists of the nilpotent element of $\mathfrak{A}$. For $z^k=0$ implies $\omega(z^k)=[\omega(z)]^k=0$, $\omega(z)=0$; conversely $\omega(z)=0$ implies $z^r=0$ by (14). It follows from Theorem 1 that the same characterization of $\mathfrak{N}$ holds for genetic algebras.

We construct an example of a train algebra which is not a genetic algebra as follows: let $\mathfrak{F}$ have characteristic two.[1] Then the square of any element z in the commutative algebra $\mathfrak{N}=(v_1,v_2,v_3)$ with multiplication table

$$v_1v_2=v_3, \qquad v_2v_3=v_1, \qquad v_3v_1=v_2, \qquad v_i^2=0 \qquad (i=1,2,3)$$

is zero. However, $\mathfrak{N}$ is not a nilpotent algebra, since $\mathfrak{N}=\mathfrak{N}^2$ $(=\mathfrak{N}\mathfrak{N})$. Let $\mathfrak{A}$ be the algebra obtained by adjoining a unity element 1 to $\mathfrak{N}$. Then x in $\mathfrak{A}$ has the form $x=\xi 1+z$, and $\mathfrak{A}$ is a train algebra since $x\rightarrow\xi$ is a weight function for $\mathfrak{A}$, while $z^2=(x-\xi 1)^2=x^2+\xi^2 1=0$ implies $x^3+\xi^2x=0$. Since $\mathfrak{N}$ is not nilpotent, it follows from Theorem 4 below that $\mathfrak{A}$ is not a genetic algebra. This example also shows that a structure theory as elementary as that in **5** below is not possible for train algebras.

A commutative baric algebra $\mathfrak{A}$ with weight function ω is called a *special train algebra* in case

(a) the kernel $\mathfrak{N}$ of ω is nilpotent, and

(b) the subalgebras $\mathfrak{N}^k$ of $\mathfrak{A}$ defined inductively by $\mathfrak{N}^1=\mathfrak{N}$, $\mathfrak{N}^k=\mathfrak{N}^{k-1}\mathfrak{N}$ for $k=2,3,\cdots$, are ideals of $\mathfrak{A}$.

---

[1] If we knew an example, over a more or less arbitrary field, of a commutative non-nilpotent algebra $\mathfrak{N}$, all of whose elements are nilpotent, we could give a more satisfying example of a train algebra which is not a genetic algebra by adjoining 1 to $\mathfrak{N}$. There are many examples in the literature of non-commutative algebras $\mathfrak{N}$ with these properties but, although it seems possible that commutative examples exist, we have not been able to construct one.

THEOREM 2. *A special train algebra* $\mathfrak{A}$ *over* $\mathfrak{F}$ *is a genetic algebra.*

Etherington has shown in [7, § 4] that over a finite extension $\mathfrak{K}$ of $\mathfrak{F}$ there exists a basis of $\mathfrak{A}_{\mathfrak{K}}$, together with scalars $\lambda_1 = 1, \lambda_2, \cdots, \lambda_n$ in $\mathfrak{K}$, such that the matrix of R_x for x in $\mathfrak{A}$ has the form

$$(18) \qquad R_x = \begin{bmatrix} \xi\lambda_1 & * & \cdots & * \\ 0 & \xi\lambda_2 & \cdots & * \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \xi\lambda_n \end{bmatrix}, \qquad \xi = \omega(x).$$

Then the characteristic function (15) of R_x has the form (16) with λ_i as in (18). Hence

$$(19) \qquad (-1)^t \gamma_t = \Sigma \lambda_{i_1} \cdots \lambda_{i_t} \qquad (t = 1, 2, \cdots, n),$$

and the γ_t in $\mathfrak{F}$ are dependent, not on x, but only on the algebra $\mathfrak{A}$. From (18) we obtain

$$f(R_{x_1}, R_{x_2}, \cdots) = \begin{bmatrix} f(\xi_1\lambda_1, \xi_2\lambda_1, \cdots) & * & \cdots & * \\ 0 & f(\xi_1\lambda_2, \xi_2\lambda_2, \cdots) & \cdots & * \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & f(\xi_1\lambda_n, \xi_2\lambda_n, \cdots) \end{bmatrix},$$

where $\xi_k = \omega(x_k)$. Then T in (12) has characteristic equation

$$\begin{aligned} |\lambda I - T| &= |(\lambda - \alpha)I - f(R_{x_1}, R_{x_2}, \cdots)| \\ &= [(\lambda - \alpha) - \eta_1][(\lambda - \alpha) - \eta_2] \cdots [(\lambda - \alpha) - \eta_n] = 0 \end{aligned}$$

where we have written

$$(20) \qquad \eta_i = f(\xi_1\lambda_i, \xi_2\lambda_i \cdots) \qquad (n = 1, 2, \cdots, n).$$

Then

$$|\lambda I - T| = (\lambda - \alpha)^n + \mu_1(\lambda - \alpha)^{n-1} + \cdots + \mu_n,$$

where $(-1)^s\mu_s$ is the elementary symmetric function of degree s in the η_i. But then by (20) the μ_s $(s = 1, 2, \cdots, n)$ are polynomials in $\xi_1, \xi_2, \cdots$, with coefficients which are symmetric functions of the λ_i. These coefficients are expressible in terms of the elementary symmetric functions (19) of the λ_i, and therefore are not dependent on the x_i. Hence $\mathfrak{A}$ is a genetic algebra.

The *copular algebra of simple mendelian inheritance* (6 below) is an example of a genetic algebra which is not a special train algebra.

4. Duplicate of a commutative algebra. Let $\mathfrak{A}$ be a commutative algebra of order n over $\mathfrak{F}$, and let $u_1, u_2, \cdots, u_n$ be a basis of $\mathfrak{A}$ with multiplication table (6). The *duplicate* $\mathfrak{A}'$ of $\mathfrak{A}$ is defined as the commutative algebra of order $\frac{1}{2}n(n+1)$ over $\mathfrak{F}$ with basal elements v_{ij} $(i \leqq j;\ i, j = 1, 2, \cdots, n)$ satisfying

$$v_{ij}v_{rs} = \Sigma_{kt}\gamma_{ijk}\gamma_{rst}v_{kt} \qquad (i \leqq j,\ r \leqq s;\ i, j, r, s, k, t = 1, \cdots, n) \tag{21}$$

where we identify $v_{tk} = v_{kt}$ for $t > k$. That is, element of $\mathfrak{A}'$ behave like quadratic forms in $\mathfrak{A}$. The definition of $\mathfrak{A}'$ is independent of the basis chosen for $\mathfrak{A}$ since $\mathfrak{A} \cong \mathfrak{A}_1$ implies $\mathfrak{A}' \cong \mathfrak{A}_1{}'$ [9, Theorem IV].

The process of duplication is important in genetics because we obtain from any gametic algebra $\mathfrak{A}$ a corresponding *zygotic algebra* $\mathfrak{A}'$ whose basis consists of the zygotic types u_iu_j $(= u_ju_i)$ obtained from the gametic types $u_1, u_2, \cdots, u_n$ in $\mathfrak{A}$. Multiplication in the zygotic algebra $\mathfrak{A}'$ is carried out as though it were being performed in $\mathfrak{A}$ according to the multiplication table (6). Writing v_{ij} for u_iu_j $(i \leqq j)$ we obtain (21), where the coefficient of v_{kt} is the probability that an individual of zygotic type u_iu_j mating with one of type u_ru_s will produce an individual of zygotic type u_ku_t. If ω is the weight function (8) of the gametic algebra $\mathfrak{A}$, then

$$\omega'\colon\ a = \Sigma_{ij}\alpha_{ij}v_{ij} \rightarrow \omega'(a) = \Sigma_{ij}\alpha_{ij}, \qquad \alpha_{ij} \text{ in } \mathfrak{F},$$

is a weight function for the zygotic algebra $\mathfrak{A}'$; $\mathfrak{A}'$ is a baric algebra. Genetical calculations involving the first filial generation may be performed in $\mathfrak{A}'$, those involving the second filial generation in the *copular algebra* $\mathfrak{A}''$ (the duplicate of $\mathfrak{A}'$), etc.

We return to the general notion of a duplicate algebra as defined by (21). There is a homomorphism H of $\mathfrak{A}'$ into $\mathfrak{A}$ defined by

$$H\colon\ v_{ij} \rightarrow v_{ij}H = u_iu_j = \Sigma_k\gamma_{ijk}u_k, \qquad (i \leqq j;\ i, j = 1, \cdots, n). \tag{22}$$

Actually H is a homomorphism of $\mathfrak{A}'$ onto the ideal $\mathfrak{A}^2$ $(= \mathfrak{A}\mathfrak{A})$ of $\mathfrak{A}$, since the u_iu_j span $\mathfrak{A}^2$. We denote the kernel of H by $\mathfrak{D}$. It is easy to see that

$$\mathfrak{D}\mathfrak{A}' = 0 \tag{23}$$

[9, Theorem II (ii)]; that is, $\mathfrak{D}$ consists of absolute divisors of zero.

Let a be an element of $\mathfrak{A}'$. We denote the corresponding right multiplication of $\mathfrak{A}'$ by R^*_a. Then any element T_* of $T(\mathfrak{A}')$ has the form

$$T_* = \alpha I_* + f(R^*_{a_1}, R^*_{a_2}, \cdots), \qquad \alpha \text{ in } \mathfrak{F},\ a_i \text{ in } \mathfrak{A}', \tag{24}$$

where I_* is the identity on $\mathfrak{A}'$.

LEMMA. *Let a commutative algebra* $\mathfrak{A}$ *of order* n *over* $\mathfrak{F}$ *have duplicate* $\mathfrak{A}'$, *and* T_* *in* $T(\mathfrak{A}')$ *have the form* (24). *Then the characteristic function of* T_* *is*

$$(25) \qquad | \lambda I_* - T_* | = (\lambda - \alpha)^{\frac{1}{2}n(n-1)} \, | \lambda I - T |$$

where T *in* $T(\mathfrak{A})$ *has the form* (12) *with* $x_i = a_i H$, *and* H *is the homomorphism* (22) *of* $\mathfrak{A}'$ *into* $\mathfrak{A}$.

Let m be the order over $\mathfrak{F}$ of the kernel $\mathfrak{Q}$ of H. Then $\mathfrak{A}' = \mathfrak{Q} + \mathfrak{D}$ for a linear subspace $\mathfrak{D}$ of $\mathfrak{A}'$ having order $p = \frac{1}{2}n(n+1) - m$ over $\mathfrak{F}$. It follows from (23) that, corresponding to this way of writing $\mathfrak{A}'$, the right multiplication $R^*{}_a$ has matrix

$$R^*{}_a = \begin{pmatrix} 0 & 0 \\ M_a & N_a \end{pmatrix}, \qquad a \text{ in } \mathfrak{A}',$$

where M_a and N_a are $p \times m$ and $p \times p$ matrices respectively. Since $\mathfrak{A}' - \mathfrak{Q} \cong \mathfrak{A}^2$ under the natural correspondence determined by H, we have $N_a = R^0{}_{aH}$, the right multiplication of $\mathfrak{A}^2$ corresponding to aH in $\mathfrak{A}^2$. Then $R^*{}_a$ has matrix

$$(26) \qquad R^*{}_a = \begin{pmatrix} 0 & 0 \\ * & R^0{}_x \end{pmatrix}, \qquad x = aH \text{ in } \mathfrak{A}^2.$$

Now $\mathfrak{A} = \mathfrak{A}^2 + \mathfrak{S}$ for a linear subspace $\mathfrak{S}$ of $\mathfrak{A}$ having order $n - p$ over $\mathfrak{F}$. Corresponding to this way of writing $\mathfrak{A}$, the matrix of R_x for x in $\mathfrak{A}^2 \subset \mathfrak{A}$ is

$$(27) \qquad R_x = \begin{pmatrix} R^0{}_x & 0 \\ * & 0 \end{pmatrix}, \qquad x \text{ in } \mathfrak{A}^2,$$

since $\mathfrak{A}^2$ is an ideal of $\mathfrak{A}$. It follows from (26) and (27) that

$$(28) \qquad f(R^*{}_{a_1}, R^*{}_{a_2}, \cdots) = \begin{pmatrix} 0 & 0 \\ * & f(R^0{}_{x_1}, R^0{}_{x_2}, \cdots) \end{pmatrix}, \qquad x_i = a_i H,$$

and

$$(29) \qquad f(R_{x_1}, R_{x_2}, \cdots) = \begin{pmatrix} f(R^0{}_{x_1}, R^0{}_{x_2}, \cdots) & 0 \\ * & 0 \end{pmatrix}, \qquad x_i \text{ in } \mathfrak{A}^2.$$

Then, denoting by I_p the (p-rowed) identity on $\mathfrak{A}^2$ and $\mathfrak{D}$, equations (28) and (29) imply that $| \lambda I_* - T_* | = (\lambda - \alpha)^m \, | (\lambda - \alpha) I_p - f(R^0{}_{x_1}, R^0{}_{x_2}, \cdots) | = (\lambda - \alpha)^{m-(n-p)} \, | \lambda I - T |$, for T_* in (24) and T in (12) with $x_i = a_i H$. Equation (25) follows immediately since $m - n + p = - n + \frac{1}{2}n(n+1) = \frac{1}{2}n(n-1)$.

We use this lemma in the proof of

THEOREM 3. *The duplicate* $\mathfrak{A}'$ *of a genetic algebra* $\mathfrak{A}$ *over* $\mathfrak{F}$ *is itself a genetic algebra.*

By definition $\mathfrak{A}'$ is commutative. If ω is the weight function of $\mathfrak{A}$, then a weight function ω' of $\mathfrak{A}'$ is defined by

$$(30) \qquad \omega'\colon\; a \to \omega'(a) = \omega(aH), \qquad a \text{ in } \mathfrak{A}',\ aH \text{ in } \mathfrak{A},$$

where H is the homomorphism (22) of $\mathfrak{A}'$ into $\mathfrak{A}$. It follows from (25) and the fact that $\mathfrak{A}$ is a genetic algebra that the characteristic function of $T_{\bullet}$ in (24) has coefficients which, insofar as they depend on the a_i, depend only on the $\omega(x_i) = \omega(a_iH) = \omega'(a_i)$; $\mathfrak{A}'$ is a genetic algebra.

The advantage of the concept of genetic algebra over special train algebra lies in Theorem 3. The most elementary algebra of genetics, the *gametic algebra of simple mendelian inheritance* (**6** below), is a special train algebra (therefore a genetic algebra by Theorem 2). Hence by Theorem 3 all algebras obtained from it by duplication are also genetic algebras. However, the copular algebra of simple mendelian inheritance, obtained by duplicating the gametic algebra twice, is not a special train algebra.

5. Structure of genetic algebras. It follows from (5) that the radical of any baric algebra is contained in the kernel $\mathfrak{N}$ of the weight function ω. We shall show that for a genetic algebra $\mathfrak{A}$ the radical *is* $\mathfrak{N}$, by showing that $\mathfrak{N}$, which we already know consists of the nilpotent elements of $\mathfrak{A}$, is actually nilpotent.

THEOREM 4. *Let* $\mathfrak{N}$ *be the kernel of the weight function* ω *of a genetic algebra* $\mathfrak{A}$ *over* $\mathfrak{F}$. *Then* $\mathfrak{N}$ *is the radical of* $\mathfrak{A}$, *and is nilpotent.*

Let T in $T(\mathfrak{A})$ have the form (12), and write $\omega(x_i) = \xi_i$. Then the characteristic function

$$|\lambda I - T| = \lambda^n + \psi_1\lambda^{n-1} + \cdots + \psi_n$$

of T has coefficients ψ_j which are polynomials in $\alpha, \xi_1, \xi_2, \cdots$, with constant terms $\psi_{j0} = 0$. For let $\alpha = 0$ and $x_i = 0$ $(i = 1, 2, \cdots)$ in (12); then $T = 0$, $|\lambda I| = \lambda^n = \lambda^n + \psi_{10}\lambda^{n-1} + \cdots + \psi_{n0}$, or $\psi_{j0} = 0$ for $j = 1, 2, \cdots, n$. It follows that λ^n is the characteristic function of any T in $T(\mathfrak{A})$ which may be written in the form (12) with

$$(31) \qquad \alpha = \omega(x_i) = 0, \qquad i = 1, 2, \cdots.$$

In this case $T^n = 0$, T is nilpotent. Now let T be in the enveloping algebra

$\mathfrak{N}^*$ of the right multiplications corresponding to elements of $\mathfrak{N}$. Then (31) is satisfied, T is nilpotent. Since $\mathfrak{N}^*$ is an associative algebra consisting of nilpotent elements, $\mathfrak{N}^*$ is nilpotent. Thus $\mathfrak{N}$ is a nilpotent ideal of $\mathfrak{A}$, and is contained in the radical $\mathfrak{R}$ of $\mathfrak{A}$. On the other hand, (5) implies that $\mathfrak{N}$ contains $\mathfrak{R}$, or $\mathfrak{R} = \mathfrak{N}$.

The classical elements of a structure theory for a linear algebra $\mathfrak{A}$ are

(i) the nature of the radical $\mathfrak{R}$, and

(ii) the nature of the simple components of the semi-simple algebra $\mathfrak{A} - \mathfrak{N}$.

For genetic algebras, (i) is answered by Theorem 4. Question (ii) is trivial by virtue of (5).

One may also ask whether or not the analogue of the so-called Wedderburn Principal Theorem holds: does $\mathfrak{A}$ contain a subalgebra $\mathfrak{S} \cong \mathfrak{A} - \mathfrak{N}$, so that $\mathfrak{A} = \mathfrak{S} + \mathfrak{N}$? It is easy to see from (5) that this question is equivalent to the following one: does $\mathfrak{A}$ contain an idempotent element e? One may readily construct an example of a genetic algebra without an idempotent, so the answer in general is negative.

However, the existence of an idempotent is significant genetically, since it represents a population in equilibrium for random mating [6, p. 138]. Etherington gives conditions for the existence of an idempotent in a commutative baric algebra [6, Theorem VI]. Clearly a genetic algebra (or even a train algebra) contains an idempotent e if and only if there is an associative subalgebra $\mathfrak{E}$ of $\mathfrak{A}$ which is not contained in $\mathfrak{N}$. For e generates such an algebra $\mathfrak{E}$, while the converse follows from the fact that any non-nilpotent associative algebra $\mathfrak{E}$ contains an idempotent.

6. Simple mendelian inheritance. Let $\mathfrak{F}$ have characteristic not two. The *gametic algebra* $\mathfrak{G}$ *of simple mendelian inheritance* is the commutative algebra $\mathfrak{G} = (u_1, u_2)$ of order 2 over $\mathfrak{F}$ with gametic multiplication table

$$u_1{}^2 = u_1, \qquad u_1u_2 = \tfrac{1}{2}u_1 + \tfrac{1}{2}u_2, \qquad u_2{}^2 = u_2. \tag{32}$$

An easy change of basis gives $\mathfrak{G} = (u, z)$ with

$$u^2 = u, \qquad uz = \tfrac{1}{2}z, \qquad z^2 = 0. \tag{33}$$

Writing $x = \xi u + \eta z$, we have weight function $\omega\colon\ x \rightarrow \omega(x) = \xi$; then $\mathfrak{N} = (z)$, $\mathfrak{N}^2 = 0$, $\mathfrak{G}$ is a special train algebra. The transformation algebra $T(\mathfrak{G})$ has order 3 over $\mathfrak{F}$, and any element T of $T(\mathfrak{G})$ may be written in the form

$$T = \alpha I + 2R_x, \qquad \alpha \text{ in } \mathfrak{F},\ x = \xi u + \eta z \text{ in } \mathfrak{G}. \tag{34}$$

The characteristic function of T in (34) is

$$(35)\qquad \lambda^2 - (2\alpha + 3\xi)\lambda + (\alpha + \xi)(\alpha + 2\xi) = [\lambda - (\alpha + 2\xi)][\lambda - (\alpha + \xi)].$$

The *zygotic algebra* $\mathfrak{Z}$ *of simple mendelian inheritance* is obtained from $\mathfrak{G}$ by duplication. For purposes of computation in genetics one would want the multiplication table obtained by duplicating (32). However, the structure of $\mathfrak{Z} = \mathfrak{G}'$ is seen more easily when the multiplication table (33) is duplicated. Write $a = uu$, $b = uz$, $c = zz$. Then $\mathfrak{Z} = (a, b, c)$ with

$$(36)\qquad a^2 = a, \qquad ab = \tfrac{1}{2}b, \qquad b^2 = \tfrac{1}{4}c, \qquad ac = bc = c^2 = 0.$$

Writing $x = \xi a + \eta b + \zeta c$, we have the weight function $\omega\colon x \rightarrow \omega(x) = \xi$. The kernel of ω is $\mathfrak{N} = (b, c)$. Then $\mathfrak{N}^2 = (c)$, $\mathfrak{N}^3 = 0$, $\mathfrak{Z}$ is a special train algebra. The transformation algebra $T(\mathfrak{Z})$ has order 6 over $\mathfrak{F}$ and any element T of $T(\mathfrak{Z})$ may be written in the form

$$(37)\qquad T = \alpha I + 2R_{x_1} + 4R_{x_2}R_{x_3}, \qquad\qquad x_i = \xi_i a + \eta_i b,\ \omega(x_i) = \xi_i.$$

The characteristic function of T in (37) is

$$(38)\qquad [\lambda - \alpha][\lambda - (\alpha + 2\xi_1 + 4\xi_2\xi_3)][\lambda - (\alpha + \xi_1 + \xi_2\xi_3)].$$

Duplication of $\mathfrak{Z}$ gives the *copular algebra* $\mathfrak{C}$ *of simple mendelian inheritance*. As before we omit the multiplication table (important from the point of view of genetics, but not structurally) which is obtained by duplicating (32) twice. Instead we write $v = aa$, $p_1 = ab$, $p_2 = bb$, $p_3 = ac$, $p_4 = bc$, $p_5 = cc$, for a, b, c in (36). Then $\mathfrak{C} = (v, p_1, \cdots, p_5)$ with

$$(39)\qquad \begin{array}{lllll} v^2 = v, & vp_1 = \tfrac{1}{2}p_1, & vp_2 = \tfrac{1}{4}p_3, & p_1{}^2 = \tfrac{1}{4}p_2, & p_1p_2 = \tfrac{1}{8}p_4, \\ p_2{}^2 = \tfrac{1}{16}p_5, & vp_j = p_ip_j = 0 & & \multicolumn{2}{l}{(i = 1, \cdots, 5;\ j = 3, 4, 5).} \end{array}$$

Writing $x = \xi v + \Sigma\pi_i p_i$, we have $\omega(x) = \xi$. The kernel of ω is $\mathfrak{N} = (p_1, \cdots, p_5)$, and $\mathfrak{N}^2 = (p_2, p_4, p_5)$, $\mathfrak{N}^3 = (p_4, p_5)$, $\mathfrak{N}^4 = 0$. Now $\mathfrak{C}$ is not a special train algebra since $\mathfrak{C}\mathfrak{N}^2$ contains $vp_2 = \tfrac{1}{4}p_3$, which is not in $\mathfrak{N}^2$; $\mathfrak{N}^2$ is not an ideal of $\mathfrak{C}$. That $\mathfrak{C}$ is a genetic algebra is guaranteed by Theorem 3.

7. Jordan algebras. A commutative algebra $\mathfrak{A}$ of order n over $\mathfrak{F}$ is called a *Jordan algebra* in case

$$x^2(xy) = x(x^2y) \qquad\qquad \text{for all } x, y \text{ in } \mathfrak{A}.$$

Let $\mathfrak{F}$ have characteristic not two; any linear subspace $\mathfrak{M}$ of $(\mathfrak{F})_m$ which is closed with respect to "quasi-multiplication"

$$(40) \qquad xy = \tfrac{1}{2}(x \cdot y + y \cdot x), \qquad x, y \text{ in } \mathfrak{M},$$

where $x \cdot y$ denotes the associative multiplication of transformations in $\mathfrak{M}$, is a *Jordan algebra of linear transformations* of order $n \leqq m^2$ over $\mathfrak{F}$ [3, p. 546].

The gametic algebra $\mathfrak{G}$ of simple mendelian inheritance is the case $n = 2$ of the genetic algebra $\mathfrak{G}_n = (u, z_2, \cdots, z_n)$ with multiplication table

$$(41) \qquad u^2 = u, \qquad uz_j = \tfrac{1}{2}z_j, \qquad z_i z_j = 0 \qquad (i, j = 2, \cdots, n).$$

In [6, § 6] Etherington has shown that any train algebra of rank 2 and order n over $\mathfrak{F}$ of characteristic not two is equivalent to $\mathfrak{G}_n$, and also has indicated that these algebras are Jordan algebras [6, p. 138, footnote]. Actually they are Jordan algebras of linear transformations. For let e_{ij} $(i, j = 1, \cdots, n)$ be the usual matric basis of $(\mathfrak{F})_n$ with matrix multiplication

$$(42) \qquad e_{ij} \cdot e_{kl} = \delta_{jk} e_{il} \qquad \text{(Kronecker delta)},$$

and let $\mathfrak{M} = (e_{11}, e_{12}, \cdots, e_{1n})$. By (40) $\mathfrak{M}$ is a Jordan algebra of linear transformations with multiplication $e_{1i}e_{1j} = \frac{1}{2}(\delta_{i1}e_{1j} + \delta_{j1}e_{1i})$ or

$$(43) \qquad e_{11}{}^2 = e_{11}, \qquad e_{11}e_{1j} = \tfrac{1}{2}e_{1j}, \qquad e_{1i}e_{1j} = 0 \qquad (i, j = 2, \cdots, n).$$

By (41) and (43) the correspondence $u \rightarrow e_{11}$, $z_i \rightarrow e_{1i}$ $(i = 2, \cdots, n)$ is an equivalence between $\mathfrak{G}_n$ and $\mathfrak{M}$.

The zygotic algebra $\mathfrak{Z}$ of simple mendelian inheritance is also a Jordan algebra of linear transformations. For let $\mathfrak{M}$ be the subspace of $(\mathfrak{F})_3$ with basal elements $a = e_{22}$, $b = e_{12} + e_{23}$, $c = 4e_{13}$. Defining multiplication in $\mathfrak{M}$ by (40) and (42), we obtain the multiplication table (36); $\mathfrak{M}$ is a Jordan algebra of linear transformations and is equivalent to $\mathfrak{Z}$.

Inasmuch as powers of a single element are associative in a Jordan algebra over $\mathfrak{F}$ of characteristic not two [3, § 5], the copular algebra $\mathfrak{C}$ of simple mendelian inheritance is not a Jordan algebra. For we see from (39) that the right power $p_1{}^4 = 0$, while $p_1{}^2 p_1{}^2 = 2^{-8}p_5 \neq 0$.

These considerations lead us to an analysis of those genetic algebras $\mathfrak{A}$ over a field $\mathfrak{F}$ of characteristic not two which are at the same time Jordan algebras. Let u be an element of weight 1 in $\mathfrak{A}$. Then, by the associativity of powers in $\mathfrak{A}$, u generates an associative subalgebra $\mathfrak{E}$ which is not contained in $\mathfrak{N}$; there is an idempotent e in $\mathfrak{A}$.

Albert has shown in [3] that the only possible characteristic roots of R_e are 0, $\frac{1}{2}$, 1, so that the equation

$$xe = \beta x, \qquad \beta \text{ in } \mathfrak{F}, \tag{44}$$

has solutions only for $\beta = 0, \frac{1}{2}, 1$. Writing $\mathfrak{A}_e(\beta)$ for the set of all x in $\mathfrak{A}$ satisfying (44), one obtains $\mathfrak{A}$ as the supplementary sum

$$\mathfrak{A} = \mathfrak{A}_e(1) + \mathfrak{A}_e(\tfrac{1}{2}) + \mathfrak{A}_e(0). \tag{45}$$

Let t_β be the dimension of the space $\mathfrak{A}_e(\beta)$ over $\mathfrak{F}$. Then $t_1 \geqq 1$ and $t_1 + t_{\frac{1}{2}} + t_0 = n$. Corresponding to a basis of $\mathfrak{A}$ in the form (45), we have the matrix of R_e in the diagonal form

$$R_e = \operatorname{diag}\{I_1, \tfrac{1}{2}I_{\frac{1}{2}}, 0\} \tag{46}$$

where I_β is the t_β-rowed identity matrix.

Since $T(\mathfrak{A})$ is a baric algebra with weight function θ defined by (10), or

$$\theta(T) = \alpha + f(\xi_1, \xi_2, \cdot \cdot \cdot), \qquad \xi_i = \omega(x_i), \tag{47}$$

for T in (12), we know that one characteristic root of T is $\theta(T)$. For $\phi(\lambda) = | \lambda I - T |$ implies that $\phi(T) = 0$ and $\theta[\phi(T)] = \phi[\theta(T)] = 0$, $\theta(T)$ is a root of $\phi(\lambda) = 0$. However, for Jordan algebras we can obtain the following stronger result.

THEOREM 5. *Let $\mathfrak{F}$ have characteristic not two, and $\mathfrak{A}$ be a genetic algebra which is a Jordan algebra over $\mathfrak{F}$. Then $\mathfrak{A}$ contains an idempotent e, and the distinct characteristic roots of T in $T(\mathfrak{A})$ in* (12) *are at most three:*

$$\alpha, \qquad \theta(T), \qquad \alpha + f(\tfrac{1}{2}\xi_1, \tfrac{1}{2}\xi_2, \cdot \cdot \cdot),$$

where $\xi_i = \omega(x_i)$, and θ is the weight function (10) *of $T(\mathfrak{A})$. The multiplicities of these roots are the orders over $\mathfrak{F}$ of $\mathfrak{A}_e(0)$, $\mathfrak{A}_e(1)$, $\mathfrak{A}_e(\frac{1}{2})$ respectively.*

For it follows from (46) that

$$f(R_{\xi_1 e}, R_{\xi_2 c}, \cdot \cdot \cdot) = \operatorname{diag}\{f(\xi_1, \xi_2, \cdot \cdot \cdot)I_1, f(\tfrac{1}{2}\xi_1, \tfrac{1}{2}\xi_2, \cdot \cdot \cdot)I_{\frac{1}{2}}, 0\}. \tag{48}$$

Then, since $\mathfrak{A}$ is a genetic algebra and $\omega(x_i) = \xi_i = \omega(\xi_i e)$, we have

$$| \lambda I - T | = | (\lambda - \alpha)I - f(R_{x_1}, R_{x_2}, \cdot \cdot \cdot) | = | (\lambda - \alpha)I - f(R_{\xi_1 e}, R_{\xi_2 e}, \cdot \cdot \cdot) | .$$

Then (48) and (47) imply that the characteristic function of T is

$$[\lambda - \alpha]^{t_0}[\lambda - \theta(T)]^{t_1}[\lambda - \{\alpha + f(\tfrac{1}{2}\xi_1, \tfrac{1}{2}\xi_2, \cdot \cdot \cdot)\}]^{t_{\frac{1}{2}}}. \tag{49}$$

The theorem follows.

In the examples we have worked out for simple mendelian inheritance, (35) and (38) illustrate the characteristic function (49).

THE INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

1. A. A. Albert, "Non-associative algebras I. Fundamental concepts and isotopy," *Annals of Mathematics*, vol. 43 (1942), pp. 685-707.
2. ———, "On Jordan algebras of linear transformations," *Transactions of the American Mathematical Society*, vol. 59 (1946), pp. 524-555.
3. ———, "A structure theory for Jordan algebras," *Annals of Mathematics*, vol. 48 (1947), pp. 546-567.
4. L. E. Dickson, *Linear algebras*, Cambridge Tract, no. 16 (1914, reprinted 1930).
5. I. M. H. Etherington, "Genetic algebras," *Proceedings of the Royal Society of Edinburgh*, vol. 59 (1939), pp. 242-258.
6. ———, "Commutative train algebras of ranks 2 and 3," *Journal of the London Mathematical Society*, vol. 15 (1940), pp. 136-149.
7. ———, "Special train algebras," *Quarterly Journal of Mathematics* (Oxford), vol. 12 (1941), pp. 1-8.
8. ———, "Non-associative algebra and the symbolism of genetics," *Proceedings of the Royal Society of Edinburgh*, vol. 61, pt. B (1941), pp. 24-42.
9. ———, "Duplication of linear algebras," *Proceedings of the Edinburgh Mathematical Society* (2), vol. 6 (1941), pp. 222-230.
10. ———, "Corrigendum: commutative train algebras of ranks 2 and 3," *Journal of the London Mathematical Society*, vol. 20 (1945), p. 238.

DIVISIBILITY PROPERTIES OF THE FOURIER COEFFICIENTS OF THE MODULAR INVARIANT $j(\tau)$.*

By JOSEPH LEHNER.

1. Introduction. Some years ago D. H. Lehmer [1] [1] gave a series of arithmetical properties of the Fourier coefficients of the modular invariant

$$j(\tau) = x^{-1} + 744 + 196,884x + \cdots, \qquad x = \exp 2\pi i\tau, I(\tau) > 0.$$

Here we derive some arithmetical properties of the coefficients of a quite different character. In fact we prove

THEOREM 1. *If*

$$j(\tau) = x^{-1} + c_0 + \sum_{\nu=1}^{\infty} c_\nu x^\nu, \tag{1.1}$$

then

$$c_{5\nu} \equiv 0 \pmod{25} \tag{1.2}$$

$$c_{7\nu} \equiv 0 \pmod{7} \tag{1.3}$$

$$c_{11\nu} \equiv 0 \pmod{11}, \qquad \nu = 1, 2, 3, \cdots. \tag{1.4}$$

The general idea behind the proof of this theorem is as follows. Consider the congruence (1.2) as an example. By applying a certain linear operator—denoted by U_5—to the right member of (1.1) one obtains the series with ν replaced by 5ν. The same operation performed on the left member yields a modular function belonging to a subgroup of the full modular group. We shall try, therefore, to express that modular function in terms of the basic function of the subgroup, which is known to have integral coefficient Fourier expansions. The identity which results is of the form

$$U_5 j(\tau) = 25(a_1\Phi(\tau) + a_2\Phi^2(\tau) + \cdots + a_s\Phi^s(\tau)), \tag{1.4$'$}$$

where Φ is the basic function, and the congruence (1.2) follows at once.

Formula (1.4)′ is an algebraic relation connecting two modular functions on the same subgroup—$U_5 j$ and Φ—i.e., it is a modular equation. Modular equations have been used by Rademacher [3], Watson [4], and the

---

* Received February 4, 1948.

[1] Numbers in square brackets refer to the bibliography at the end of the paper.

author [2] to discuss the famous congruences of Ramanujan concerning the partition function. Ramanujan himself proposed certain identities from which his congruences would follow as immediate consequences in the same way as (1. 2) follows from (1. 4)′. Later these identities were proved using the methods of modular function theory. In particular, the proof of Theorem 1 as outlined above is modeled after Rademacher's proof of the Ramanujan congruences [3].

In the work on the partition congruences an essential inelegance occurs because of the fact that the partition function is not quite the coefficient function of the Fourier expansion of any modular function. The theory of modular functions must therefore be distorted somewhat so that an application to the partition problem is possible. The resulting awkwardness is justified, of course, by the importance of the problem. The author feels, however, that from the standpoint of the theory of *modular functions,* the developments of this paper place the theory of modular identities in a simple and natural setting.

The congruences (1. 2)-(1. 4) can be extended to powers of their respective primes. In **4-6** we prove

THEOREM 2. *With the* c_ν *of* (1. 1), we have

$$c_\nu \equiv 0 \pmod{5^{\alpha+1}} \quad \text{if } \nu \equiv 0 \pmod{5^\alpha} \tag{1.5}$$

$$c_\nu \equiv 0 \pmod{7^{\alpha}} \quad \text{if } \nu \equiv 0 \pmod{7^\alpha} \tag{1.6}$$

$$c_\nu \equiv 0 \pmod{11^{2}} \quad \text{if } \nu \equiv 0 \pmod{11^2}, \tag{1.7}$$

for $\nu = 1, 2, 3, \cdots$, *and* $\alpha = 1, 2, 3, \cdots$.

The difficulty which prevents ready extension of (1. 7) to $\alpha > 2$ will appear in due course (**4**).

Note that no statement concerning c_0 is made in Theorems 1 and 2. Obviously any $j(\tau) +$ const. has the same modular properties as $j(\tau)$ itself, so that c_0 is quite arbitrary.

2. Congruences for the moduli 5 and 7.

Let us recall first that an *entire modular function on the group* Γ is an analytic function $f(\tau)$, which is

(a) regular in the upper half-plane $I(\tau) > 0$,

(b) meromorphic with respect to an appropriate uniformizing variable when τ assumes real rational values or when $\tau = \infty$,

(c) invariant under Γ, i.e., $f(V\tau) = f(\tau)$ for every modular substitution $V\tau = (a\tau + b)/(c\tau + d)$ belonging to Γ.

We assume Γ to be a subgroup of finite index in the full modular group $\Gamma(1)$. The subgroup of interest to us is $\Gamma_0(p)$, defined by $c \equiv 0 \pmod{p}$, p being a prime > 3. $j(\tau)$, of course, belongs to the full group.

In order to get a hold on the coefficients of $j(\tau)$ whose indices are divisible by p ($p = 5, 7, 11$ in the cases of interest) we introduce the operator [2]

$$U_p f(\tau) = p^{-1} \sum_{\lambda=0}^{p-1} f((\tau + \lambda)/p). \tag{2.1}$$

First we note ([2], p. 499) that if $f(\tau) = \sum_{\nu=s}^{\infty} a_\nu x^\nu$, then

$$U_p f = \sum_{\mu=t}^{\infty} a_{p\mu} x^\mu, \qquad t = -[-s/p], \tag{2.2}$$

where, throughout this paper, $x = \exp 2\pi i\tau$. For the left member is

$$p^{-1} \sum_{\lambda=0}^{p-1} \sum_{\nu=s}^{\infty} a_\nu \exp\{2\pi i(\tau + \lambda)\nu/p\} = p^{-1} \sum_{\nu=s}^{\infty} a_\nu \exp(2\pi i\nu\tau/p) \sum_{\lambda=0}^{p-1} \exp(2\pi i\lambda\nu/p)$$

$$= \sum_{p \mid \nu} a_\nu \exp(2\pi i\tau\nu/p) = \sum_{\mu=t}^{\infty} a_{p\mu} \exp(2\pi i\mu\tau).$$

Thus in order to prove a property of the a_ν of f for which $p \mid \nu$ we may prove the same property for *all* of the a_ν of $U_p f$.

Next we prove a

LEMMA. *$U_p j$ is a modular function on $\Gamma_0(p)$.*

This lemma may be established by the same reasoning employed in the proof of Theorem 4 of [2] (pp. 499-501). That $U_p j$ is regular in the upper half-plane is obvious. The subgroup $\Gamma_0(p)$ has only two parabolic vertices when p is prime, say, $\tau = i\infty$, $\tau' = 0$. From (2.2) we note that at $\tau = i\infty$ we have the expansion

$$U_p j(\tau) = c_0 + c_p x + c_{2p} x^2 + \cdots. \tag{2.3}$$

At $\tau' = 0$ we write $\tau' = -1/p\tau$ and consider $\tau \to i\infty$. From (8.81) of [2] we then find

$$pU_p j(\tau') = pU_p j(p\tau) + j(-1/p^2\tau) - j(\tau). \tag{2.3$'$}$$

Using $j(-1/\tau) = j(\tau)$ and (2.2) we obtain

$$pU_p j(\tau') = \text{const.} + \cdots + x^{-p^2} + \cdots - x^{-1} + \cdots = x^{-p^2} + \cdots. \tag{2.4}$$

Thus $U_p j$ is regular at $\tau = i\infty$ and has a pole of order p^2 (in $x = \exp 2\pi i\tau$) at $\tau' = 0$.

To establish the invariance of U_pj on $\Gamma_0(p)$ we make use of Lemma 1 of [2], p. 501. We have

$$\begin{aligned} U_pj(V\tau) &= p^{-1}\sum_{\lambda=0}^{p-1} j(T_\lambda V\tau), && T_\lambda\tau = (\tau+\lambda)/p,\ V \mathrel{\varepsilon} \Gamma_0(p), \\ &= p^{-1}\sum_{\mu=0}^{p-1} j(W_\mu T_\mu\tau), && W_\mu \mathrel{\varepsilon} \Gamma_0(p^2), \\ &= p^{-1}\sum_{\mu} j(T_\mu\tau) = U_pj(\tau). \end{aligned}$$

Since U_pj is a modular function on $\Gamma_0(p)$ we can express it as a rational function (in fact, a polynomial) in the basic functions of that group. When $\Gamma_0(p)$ is of genus zero[2] there is a univalent function the powers of which in linear combination generate all the entire modular functions of the group. This situation obtains when $p = 5, 7, 13$, and the univalent basic functions are

$$(2.5)\qquad \Phi_{p,r}(\tau) = \Phi(\tau) = \{\eta(p\tau)/\eta(\tau)\}^r,$$

where $r(p-1) \equiv 0$, (mod 24) and r is minimal positive. Here $\eta(\tau)$ is the Dedekind function

$$(2.5)'\qquad \eta(\tau) = \exp(\pi i\tau/12)\prod_1^\infty (1-x^m).$$

Φ is obviously regular at $\tau = i\infty$, and near $\tau' = 0$ we have the integral coefficient expansion, [2], (8.83),

$$(2.6)\qquad p^{r/2}\Phi_{p,r}(\tau') = \Phi_{p,r}^{-1}(\tau) = x^{-1}(1+O(x)), \qquad \tau' = -1/p\tau.$$

It follows in view of (2.4) that by proper choice of integers $C_1, C_2, \cdots, C_k$ we can make sure that the function

$$pU_pj(\tau') - \{C_kp^{rk/2}\Phi^k(\tau') + \cdots + C_2p^r\Phi^2(\tau') + C_1p^{r/2}\Phi(\tau')\}, \qquad k = p^2,$$

does not have a pole at $\tau' = 0$.

At $\tau = i\infty$, U_pj and Φ are both regular. Hence, with a suitable constant C_0, we conclude that the function

$$\delta(\tau) = U_pj(\tau) - p^{-1}\{C_1p^{r/2}\Phi(\tau) + \cdots + C_kp^{rk/2}\Phi^k(\tau)\} - C_0$$

is a modular function of $\Gamma_0(p)$, which is regular in the interior of the fundamental region and at both parabolic vertices $(0, i\infty)$, and has a zero at

---

[2] By the genus of a group we mean the topological genus of the closed manifold obtained by identifying the edges of its fundamental region which are congruent under the substitutions of the group.

$\tau = i\infty$. By a well-known theorem, δ vanishes identically. Thus we have established the *identity*

$$U_p j(\tau) = C_0 + p^{r/2-1}\{C_1\Phi(\tau) + \cdots + C_k p^{r(k-1)/2}\Phi^k(\tau)\}. \tag{2.7}$$

From (2.2) and the last equation we write

$$\sum_{\mu=0}^{\infty} c_{p\mu} x^{\mu} = C_0 + p^{r/2-1} \sum_{\nu=1}^{\infty} b_\nu x^\nu, \tag{2.8}$$

where the b_ν are integers.

Now for $p = 5$, $r = 6$; for $p = 7$, $r = 4$; for $p = 13$, $r = 2$. Hence, when p is 5, 7, the congruences (1.2), (1.3) follow trivially from (2.8). When $p = 13$, however, $r/2 - 1 = 0$, so no obvious congruence follows.

3. Congruences for the modulus 11. When we turn to $p = 11$, we face a new situation, for $\Gamma_0(11)$ is no longer of genus zero. Therefore, there cannot be a univalent function on this subgroup, but it is possible to define a pair of independent 2-valent functions, which together constitute a rational basis ([2], pp. 501-505). These functions have integral coefficient expansions, of which the leading terms are given below. At $\tau = i\infty$, $x = \exp 2\pi i\tau$,

$$\begin{aligned} A(\tau) &= x^{-1} + 6 + 17x + \cdots, \\ C(\tau) &= x + 5x^2 + \cdots. \end{aligned} \tag{3.1}$$

At $\tau' = 0$, we put $\tau' = -1/11\tau$ and have

$$\begin{aligned} A(\tau') &= A(\tau) = x^{-1} + \cdots, \\ 11^2 C(\tau') &= x^{-2} + 2x^{-1} + \cdots. \end{aligned} \tag{3.2}$$

We wish to represent $11U_{11}j$ as a polynomial in A and C. The procedure to be employed is explained in [2], p. 505. At $\tau = i\infty$, $11U_{11}j$ is regular, as we saw in (2.3), but A has a pole. However, C and 11^2AC are regular. At $\tau' = 0$, A has a simple pole, and 11^2C a double pole. Hence, we may remove the principal part of $11U_{11}j$ at $\tau' = 0$ by means of a polynomial with terms like $(11^2C)^m$, $A \cdot (11^2C)^m$, the former being used for poles of even order, the latter for those of odd order greater than one. These terms will remain regular at $\tau' = i\infty$ because of the regularity of C and 11^2AC there. Moreover, such terms all have integral coefficient expansions. We have not yet considered the term in x^{-1} in the principal part of $11U_{11}j$ at $\tau' = 0$. However, this term must drop out since otherwise we could construct an entire modular function having but one simple pole on $\Gamma_0(11)$, a subgroup of genus one. The above considerations lead to an identity of the form

$$(3.3)\quad U_{11}j(\tau) = D + 11\{D_1C(\tau) + D'_1A(\tau)C(\tau) + D_2 11^2C^2(\tau) + D'_2 11^2 A(\tau)C^2(\tau) + \cdots + D_t 11^{2t-2}C^t(\tau) + D'_t 11^{2t-2}A(\tau)C^t(\tau)\},$$

where $D_1, \cdots, D_t;\ D'_1, \cdots, D'_t$, are integers. The expression in braces, when expanded in powers of x, clearly has integral coefficients. The congruence (1.4) follows from this identity.

4. Congruence for the modulus 11^2. It is convenient at this point to indicate how the identity (3.3) can be extended to the modulus 11^2. This is accomplished by subjecting both members of the identity to the operator U_{11}. On the left we obtain by (2.2)

$$U_{11}{}^2 j(\tau) = U_{11}\{U_{11}j(\tau)\} = \sum_{\mu=0}^{\infty} c_{121\mu}x^{\mu}.$$

In the right member of (3.3) every term contains the factor 11 to at least the second power except possibly D, in which we are not interested, and D_1, D'_1. If we now apply the operator U_{11}, which merely replaces each coefficient a_μ of a power series by $a_{11\mu}$, and take congruences to the modulus 11^2, we evidently obtain

$$(4.1)\quad U_{11}{}^2 j(\tau) \equiv D + 11D_1U_{11}C(\tau) + 11D'_1U_{11}A(\tau)C(\tau) \pmod{11^2}$$
$$\equiv D' + 11D_1U_{11}C(\tau) + 11D'_1U_{11}\{A(\tau)C(\tau) - 1\} \pmod{11^2}.$$

It is proved in [2] ((9.1), p. 512 and Lemma 4, p. 504) that

$$U_{11}C(\tau) \equiv 0 \pmod{11}$$
$$A(\tau)C(\tau) - 1 \equiv 0 \pmod{11}.$$

Hence

$$(4.2)\qquad U_{11}{}^2 j(\tau) = \sum_{0}^{\infty} c_{121\mu}x^{\mu} \equiv D' \pmod{11^2},$$

and the congruence (1.7) is proved.

It is clear that the foregoing treatment cannot be extended to higher powers of 11. What is needed is a theorem about $U_{11}C^k(\tau)$, $U_{11}A(\tau)C^k(\tau)$, and the higher iterates $U_{11}{}^2, U_{11}{}^3, \cdots$. We would like this theorem to say that $U_{11}C^k$ is a polynomial of the same form as the expression in braces in (3.3) but with an additional factor of 11.

When $p = 5$ or 7 it is possible to prove theorems of this general character, as will be shown in the two following sections. The reason is this: when $p = 5$ and 7, $\Gamma_0(p)$ is of genus zero. There is a univalent function Φ on $\Gamma_0(p)$. This function satisfies a *modular equation,* i. e., an algebraic relation between $\Phi(\tau)$ and $\Phi(\tau/p)$. The iterates $U_p\Phi^k, U_p^2\Phi^k, \cdots$, can be calculated by applying Newton's formula for the sums of powers of the roots of an algebraic equation to the modular equation, a device first used by Watson [4]. Since $U_{11}j, U_{11}^2j, \cdots$, are expressible as polynomials in Φ, the application of Newton's formula yields theorems of the type mentioned above. Thus it is possible to iterate the identity (2. 7) indefinitely to obtain identities for U_p^mj, $m = 1, 2, 3, \cdots$. This will be carried out in the next two sections.

This approach unfortunately breaks down when $p = 11$, due to the nonexistence of modular equations of the required type for the functions C^k and AC^k.

5. Congruences with powers of 5 as moduli. Our next step will be to generalize the identity (2. 7) so that it involves powers of p in the indices of the coefficients $\{c_\mu\}$. This is accomplished by repeated applications of the operator U_p. We denote the iterate of U_p by U_p^2, etc.:

$$(5.1) \qquad U_p^{m+1}j(\tau) = U_p\{U_p^mj(\tau)\}, \qquad m = 1, 2, 3, \cdots.$$

In the first place it is clear that

$$U_p^2j = U_p^2\{x^{-1} + \sum_{\mu=0}^{\infty} c_\mu x^\mu\} = \sum_{\mu=0}^{\infty} c_{q\mu}x^\mu, \qquad q = p^2,$$

and in general,

$$(5.2) \qquad U_p^mj = \sum_{\mu=0}^{\infty} c_{r\mu}x^\mu, \qquad r = p^m.$$

On the other hand, iteration of (2. 7) gives

$$(5.3) \quad U_p^2j = C_0 + p^{r/2-1}U_p\{C_1\Phi + C_2p^{r/2}\Phi^2 + \cdots + C_kp^{r(k-1)/2}\Phi^k\}, \qquad k = p^2.$$

In order that the identity (5. 3) shall give rise to a new congruence to the modulus p^2, it is necessary that $U_p\Phi, \cdots, p^{r(k-1)/2}U_p\Phi^k$ should be divisible by p, at least. However, we also want expressions for $U_p\Phi, \cdots$, which are in a form suitable for iteration of U_p, since we are interested in identities and congruences to the moduli $p^3, p^4, \cdots$.

The simplest way of meeting both requirements is to prove that the polynomial in brackets in (5. 3) reproduces itself under the operation U_p

but picks up a factor p^λ, $\lambda \geq 1$. This would be true for the polynomial if it were true for each power of Φ separately. That is, we are led to consider the statement

$$(5.4)\quad p^{r(l-1)/2}U_p\Phi^l = p^\lambda\{C_{1,l}\Phi + C_{2,l}p^{r/2}\Phi^2 + C_{3,l}p^r\Phi^3 + \cdots + C_{u,l}p^{r/2(u-1)}\Phi^u\},$$
$$\lambda \geq 1,\ l = 1, 2, \cdots, p^2,\ u = u_l,$$

where $C_{1,l}, \cdots, C_{u,l}$ are integers.

This statement, however is not true. It breaks down even for $l = 1$ as we shall notice later (cf. (5. 11), (6. 11)).

We therefore proceed as follows. Rewrite (2. 7) in the form

$$(5.5)\quad U_pj = B_0 + p^{r/2-1}(B_1\Phi + B_2p^2\Phi^2 + B_3p^3\Phi^3 + \cdots + B_kp^k\Phi^k),$$

the B's being integers. We shall prove that under application of U_p the above polynomial does reproduce itself and at the same time picks up a factor p^λ:

$$U_p\Phi = p^\lambda\{B_{1,1}\Phi + B_{2,1}p^2\Phi^2 + \cdots + B_{v,l}p^v\Phi^v\},$$
$$(5.6)\quad p^lU_p\Phi^l = p^\lambda\{B_{1,l}\Phi + B_{2,l}p^2\Phi^2 + \cdots + B_{w,l}p^w\Phi^w\},$$
$$\lambda \geq 1,\ l \geq 2,\ w = w_l.$$

For this purpose we make use of the modular equation connecting $\Phi(\tau)$ and $\Phi(p\tau)$. These equations have been given in the cases $p = 5, 7$ by Rademacher [3], (11. 5) and (12. 5).

At this point we separate the cases $p = 5$ and $p = 7$, and consider the former only in this section. Putting

$$(5:7)\qquad Z = \Phi(\tau) = \{\eta(5\tau)/\eta(\tau)\}^6, \qquad W = 5^3\Phi(\tau/5),$$

we have from [3],

$$(5.8)\qquad W^5 + \sum_{j=1}^{5}(-1)^jp_jW^{5-j} = 0,$$

where

$$(-1)^{j+1}p_j = 5^5\sum_{l=j}^{5}b_lZ^{l-j+1},$$

$$(5.9)\qquad b_1 = 63 \qquad b_3 = 63.5^5 \qquad b_5 = 5^{10}.$$
$$b_2 = 52.5^3 \qquad b_4 = 6.5^8$$

(5. 8) results from replacing X and Y in Rademacher's equation by

$$Z(\tau) = Y(\tau/5), \qquad W(\tau) = X^{-1}(\tau/5).$$

The conjugates in W of (5.8) are clearly

$$W_\lambda = 5^3\Phi((\tau+\lambda)/5), \qquad \lambda = 0, 1, 2, 3, 4, \tag{5.10}$$

since replacing τ by $\tau + 1$ leaves Z unaltered, by (2.5), (2.5)′. Hence, for the sum of the conjugates, we have

$$\begin{aligned} U_5\Phi(\tau) &= 5^{-1}\sum_{\lambda=0}^{4}\Phi(\tau+\lambda/5) \\ &= 5^{-4}\sum_{\lambda=0}^{4} W_\lambda = 5^{-4}p_1 = 5\sum_{l=1}^{5} b_l Z^l. \end{aligned} \tag{5.11}$$

In order to facilitate further computation, we introduce the symbols

$$P = a_1 5\Phi + a_2 5^2\Phi^2 + \cdots + a_s 5^s\Phi^s,$$
$$Q = b_1\Phi + b_2 5^2\Phi^2 + \cdots + b_t 5^t\Phi^t,$$

where the a's, b's, s, and t are integers ($s \geq 1$, $t > 1$). P will denote a polynomial of this type, not necessarily the same one at each appearance; likewise for Q.

Obviously the sum and product of two P's is again a P; this is not necessarily true for two Q's. Moreover, every P is a Q.

In terms of this notation we can rewrite (5.6) as

$$\begin{aligned} U_5\Phi &= 5^\lambda Q, \\ 5^l U_5\Phi^l &= 5^\lambda Q, \end{aligned} \qquad \lambda \geq 1,\ l \geq 2. \tag{5.12}$$

On the other hand, we have from (5.9)

$$p_1 = 5^5 Q; \tag{5.12′}$$

then from (5.11)

$$U_5\Phi = 5Q. \tag{5.13}$$

Comparison of (5.12) and (5.13) shows that we must take $\lambda = 1$. The first part of (5.12) is therefore proved, but we must still show

$$5^l U_5\Phi^l = 5Q, \qquad l = 2, 3, \cdots. \tag{5.14}$$

Let

$$S_k = \sum_{\lambda=0}^{4} W_\lambda^k = 5^{3k}\sum_{\lambda=0}^{4}\Phi^k((\tau+\lambda)/5), \qquad k \geq 1. \tag{5.15}$$

Then

$$U_5\Phi^l = 5^{-1}\sum_{\lambda=0}^{4}\Phi^l((\tau+\lambda)/5) = 5^{-3l-1}\sum_{\lambda=0}^{4}W_\lambda{}^l$$
$$= 5^{-3l-1}S_l,$$

so (5.14) is equivalent to

$$(5.16) \qquad S_l = 5^{2l+2}Q, \qquad l = 2, 3, \cdots.$$

This relation can be proved by the aid of Newton's formula for the sums of powers of the roots of an algebraic equation. With the notation of (5.8) and (5.15), Newton's formula is

$$(5.17) \qquad S_k = \sum_{j=1}^{k}(-1)^{j+1}p_jS_{k-j}, \qquad k = 1, 2, 3, \cdots,$$

with the usual conventions

$$p_6 = p_7 = \cdots = 0, \qquad S_0 = k.$$

Although (5.16) is required only for $l \geq 2$, we shall find it convenient to establish it for all positive integral l. In view of the recursive nature of (5.17) we shall evidently proceed by induction. For $l = 1$, (5.16) is contained in (5.12)′, since, from (5.17), $S_1 = p_1$.

In the process of establishing the inductive step in (5.17), we shall have to multiply the polynomials p_j and S_{k-j} and add the products. This operation will be facilitated if these polynomials are elements of a ring. For this purpose we choose the P polynomials defined above.

We assume in the induction that

$$(5.18) \qquad S_n = 5^{2n+2}Q, \qquad n = 1, 2, \cdots, l-1.$$

Since evidently

$$5^tQ = 5^{t-1}P, \qquad t \geq 1,$$

we have

$$(5.18)' \qquad S_k = 5^{2n+1}P, \qquad n = 1, 2, \cdots, l-1.$$

Also, let

$$(5.19) \qquad p_n = 5^{\alpha_n}P, \qquad 1 \leq n \leq 5,$$

$$\alpha_1 = 4 \qquad \alpha_3 = 9 \qquad \alpha_5 = 14.$$
$$\alpha_2 = 7 \qquad \alpha_4 = 12$$

The α_k are obtained from (5.9).

If $l \leq 5$ we rewrite (5.17) as

$$(5.20) \qquad S_l = \sum_{j=1}^{l-1} (-1)^{j+1} p_j S_{l-j} - (-1)^l l p_l .$$

Then applying (5.18)′, (5.19) to (5.20), we get

$$S_l = \sum_{1}^{l-1} 5^{\alpha_j + 2(l-j)+1} P + 5^{\alpha_l} P.$$

From (5.19) we see that $\alpha_j \geq 2j + 2$; hence,

$$S_l = \sum_{1}^{l-1} 5^{2j+2+2(l-j)+1} P + 5^{2l+2} P = 5^{2l+2} P = 5^{2l+2} Q,$$

proving (5.16) when $l \leq 5$.

For $l \geq 5$, substitute (5.18)′, (5.19) in (5.17) and obtain

$$S_l = \sum_{j=1}^{l} 5^{2j+2+2(l-j)+1} P = 5^{2l+3} P = 5^{2l+2} Q.$$

This proves (5.16) for all $l \geq 1$ and thereby (5.14) (for $l \geq 2$). In turn this establishes (5.12) with $\lambda = 1$, which is equivalent to (5.16) ($p = 5, \lambda = 1$). If we operate on (5.5) with U_5 and make use of (5.6) we are enabled to write the *identity*

$$(5.21) \quad U_5{}^2 j(\tau) = A_0 + 5^3 (A_1 \Phi(\tau) + A_2 5^2 \Phi^2(\tau) + \cdots + A_t 5^t \Phi^t(\tau)),$$

where the A's and t are integers.

An immediate consequence of (5.21) is the congruence (1.5) of Theorem 2 for $\alpha = 2$.

We can now continue in the same way to apply the operator U_5 to both sides of (5.21). Using (5.6) (with $\lambda = 1$) one readily establishes by induction the identities

$$(5.22) \quad U_5{}^m j(\tau) = A_{0,m} + 5^{m+1} \{ A_{1,m} \Phi(\tau) + A_{2,m} 5^2 \Phi^2(\tau) + \cdots + A_{t,m} 5^t \Phi^t(\tau) \}, \qquad m = 1, 2, 3, \cdots,$$

where $A_{0,m}, A_{1,m}, \cdots, A_{t,m}$ are integers. Equation (1.5) of Theorem 2 follows at once.

6. Congruences with powers of 7 as moduli. The developments of the preceding section can be extended without difficulty to the case $p = 7$. We shall therefore merely list the relevant relations using the same equation number as in **5**, i. e., $(5.13) \to (6.13)$, etc. The proofs all go through in the same way.

$$U_7 j = B_0 + 7(B_1\Phi + B_2 7^2\Phi^2 + \cdots + B_k 7^k \Phi^k) \tag{6.5}$$

$$Z = \Phi(\tau) = \{\eta(7\tau)/\eta(\tau)\}^4, \qquad W = 7^2\Phi(\tau/7) \tag{6.7}$$

$$W^7 + \sum_{j=1}^{7} (-1)^j p_j W^{7-j} = 0 \tag{6.8}$$

$$(-1)^{j+1} p_j = 7^4 \sum_{l=j}^{7} b_l Z^{l-j+1} \tag{6.9}$$

$$\begin{array}{lll} b_1 = 82 & b_4 = 272 \cdot 7^5 & b_7 = 7^{10} \\ b_2 = 176 \cdot 7^2 & b_5 = 46 \cdot 7^7 & \\ b_3 = 845 \cdot 7^3 & b_6 = 4 \cdot 7^9 & \end{array}$$

$$U_7\Phi = 7^{-3} \sum_{\lambda=0}^{6} W_\lambda = 7^{-3} p_1 = 7 \sum_{l=1}^{7} b_l Z^l \tag{6.11}$$

$$U_7\Phi = 7Q \tag{6.13}$$

$$S_l = 7^{l+2}Q, \qquad l = 2, 3, \cdots \tag{6.16}$$

$$S_k = \sum_{j=1}^{k} (-1)^{j+1} p_j S_{k-j}, \qquad k = 1, 2, \cdots, \tag{6.17}$$

$$p_8 = p_9 = \cdots = 0, \qquad S_0 = k$$

$$S_n = 7^{n+2}Q, \qquad n = 1, 2, \cdots, l-1 \tag{6.18}$$

$$p_n = 7^{\alpha_n}P, \qquad 1 \leq n \leq 7, \tag{6.19}$$

$$\begin{array}{lll} \alpha_1 = 3 & \alpha_4 = 8 & \alpha_7 = 13 \\ \alpha_2 = 5 & \alpha_5 = 10 & \\ \alpha_3 = 6 & \alpha_6 = 12 & \end{array}$$

$$\alpha_j \geq j + 2$$

$$S_l = \sum_{j=1}^{l-1} 7^{\alpha_j + l - j + 1}P + 7^{l+2}P$$

$$= \sum_{1}^{l-1} 7^{j+2+l-j+1}P + 7^{l+2}P = 7^{l+2}P = 7^{l+2}Q$$

$$S_l = \sum_{1}^{l} 7^{\alpha_j + l - j + 1}P = 7^{l+3}P = 7^{l+2}Q.$$

The identities which result from these considerations are

$$(6.22)\quad U_7{}^m j(\tau) = A_{0,m} + 7^m\{A_{1,m}\Phi(\tau) + A_{2,m}7^2\Phi^2(\tau) + \cdots + A_{t,m}7^t\Phi^t(\tau)\}, \qquad m = 1, 2, 3, \cdots,$$

the A's being integers. This proves equation (1.6) of Theorem 2 and completes the proof of that theorem.

Added in Proof: Equation (1.7) of Theorem 2 is valid for the modulus 11^3, in other words, the congruence

$$c_\nu \equiv 0 \pmod{11^3} \text{ if } \nu \equiv 0 \pmod{11^3}, \qquad \nu = 1, 2, 3, \cdots,$$

is correct. The theorems required for the proof of this result are contained in a paper by the author, "Proof of Ramanujan's Partition Congruence for the Modulus 11^3," submitted to the *Bulletin of the American Mathematical Society* on Sept. 13, 1948. The required formulae are listed as equations (A) and (B) near the end of § 2 of that paper.

New York City.

BIBLIOGRAPHY.

1. D. H. Lehmer, "Properties of the coefficients of the modular invariant $J(\tau)$," *American Journal of Mathematics*, vol. 64 (1942), pp. 488-502.
2. J. Lehner, "Ramanujan identities involving the partition function for the moduli 11^a," *American Journal of Mathematics*, vol. 65 (1943), pp. 492-520.
3. H. Rademacher, "The Ramanujan identities under modular substitutions," *Transactions of the American Mathematical Society*, vol. 51 (1942), pp. 609-636.
4. G. N. Watson, "Ramanujans Vermutung über Zerfällungsanzahlen," *Journal für die reine und angewandte Mathematik*, vol. 179 (1938), pp. 97-128.

LIE AND JORDAN TRIPLE SYSTEMS.*

By Nathan Jacobson.

The present paper is devoted to a study of subspaces of an associative algebra that are closed relative to the ternary operation $[[a,b],c]$ where $[a,b]=ab-ba$. Such systems—called Lie triple systems—arise in a natural way in the study of Jordan algebras and of Jordan triple systems. The latter are defined to be subspaces of an associative algebra that are closed relative to $\{\{a,b\},c\}$ where $\{a,b\}=ab+ba$. In the first part of this paper we consider some general properties of such systems. The second half of our paper is concerned with the study of certain particular Lie and Jordan triple systems that have arisen in quantum mechanics. These systems have a basis $g_1, g_2, \cdots, g_n$ and multiplication tables, respectively

$$[[g_i, g_j], g_k] = \delta_{ki} g_j - \delta_{kj} g_i$$

$$g_i g_j g_k + g_k g_j g_i = -\delta_{ij} g_k - \delta_{kj} g_i.$$

The latter relations have been introduced by Duffin [1] and by Kemmer [2] in the study of meson fields and there is an extensive literature on the representation theory of such systems. In this paper we consider an extension of this theory.

I. Elementary Properties.

1. Lie and Jordan triple systems. We recall that a subspace $\mathfrak{L}$ of an associative algebra $\mathfrak{A}$ is a *Lie algebra* if $\mathfrak{L}$ is closed relative to the commutator composition $[a,b]\equiv ab-ba$. Similarly, a subspace $\mathfrak{J}$ is a *special Jordan algebra* if it is closed relative to the Jordan product $\{a,b\}=ab+ba$. A typical example of a Lie algebra is the set of skew symmetric matrices of n rows and columns while an example of a Jordan algebra is the set of symmetric matrices. We now define a *Lie triple system* (L. t. s.) $\mathfrak{L}_3$ as a subspace of an associative algebra that is closed under the ternary composition $[[a,b],c]$. Similarly, a subspace $\mathfrak{J}_3$ of an associative algebra will be called a *Jordan triple system* (J. t. s.) if it is closed relative to $\{\{a,b\},c\}$. Of course, any Lie (Jordan) algebra is a Lie (Jordan) triple system. On the

---

* Received February 22, 1948.

[1] Duffin [6].

[2] Kemmer [10].

other hand, the set of skew symmetric matrices is an example of a J. t. s. that is not a Jordan algebra.

We shall obtain next two characterizations of Jordan triple systems. For this purpose we introduce the composition $(abc) = abc + cba$ and we note the following relations:

$$(1) \qquad [[a, b], c] = (abc) - (bac)$$

$$(2) \qquad \{\{a, b\}, c\} = (abc) + (bac)$$

$$(3) \qquad \{\{a, b\}, c\} - \{a, \{b, c\}\} = [[c, a], b]$$

$$(4) \qquad \{\{a, b\}, c\} + \{\{b, c\}, a\} - \{\{c, a\}, b\} = 2(abc)$$

$$(5) \qquad \{\{a, a\}, a\} = 4a^3$$

$$(6) \qquad (a + b + c)^3 - (a + b)^3 - (a + c)^3 - (b + c)^3 + a^3 + b^3 + c^3 = (abc) + (bac) + (acb).$$

THEOREM 1. *A subspace* $\mathfrak{J}_3$ *of an associative algebra is a Jordan triple system if and only if either one of the following conditions holds*: (1) $\mathfrak{J}_3$ *is closed relative to* $(abc) = abc + cba$, (2) $\mathfrak{J}_3$ *is a Lie triple system closed relative to cubes. It is assumed that the characteristic of the underlying field is* $\neq 2, 3$.

Proof. If $\mathfrak{J}_3$ is a J. t. s., then by (4) $\mathfrak{J}_3$ is closed relative to (abc). Conversely by (2) a subspace closed relative to (abc) is a J. t. s. Also (1) and (5) shows that a J. t. s. is a L. t. s. closed relative to cubes. Finally, let $\mathfrak{J}_3$ be a L. t. s. closed relative to cubes. Then by (6)

$$(abc) + (bac) + (acb) \in \mathfrak{J}_3$$

and by (1), $(abc) - (bac)$ and $(abc) - (acb) \in \mathfrak{J}_3$. Hence $3(abc) \in \mathfrak{J}_3$ and $(abc) \in \mathfrak{J}_3$. Thus $\mathfrak{J}$ is a Jordan triple system.

COROLLARY. *Any special Jordan algebra is a Lie triple system.*

A part of our result is that any J. t. s. is a L. t. s. Also it is clear from (1) or from (3) that if two J. t. s. are isomorphic then they are also isomorphic as L. t. s. Hence we may speak of *the Lie triple system* of a J. t. s. In particular, we can associate a uniquely determined L. t. s. with any special Jordan algebra. Because of the relation (3) we shall also call this L. t. s. *the associator* L. t. s. of the special Jordan algebra.

2. Examples. We consider now some important instances of triple systems.

1) *Meson systems.* Let $\mathfrak{M}$ be the subspace of the algebra Φ_{n+1} of $(n+1) \times (n+1)$ matrices over the field Φ spanned by the particular matrices $g_i = e_{i,n+1} - e_{n+1,i}$, $i = 1, 2, \cdots, n$. $\mathfrak{M}$ is a J. t. s. since

$$(g_i g_j g_k) = -\delta_{ij} g_k - \delta_{kj} g_i. \tag{7}$$

For the L. t. s. of $\mathfrak{M}$ we have the multiplication table

$$[[g_i, g_j], g_k] = \delta_{ki} g_j - \delta_{kj} g_i. \tag{8}$$

The systems $\mathfrak{M}$ have been studied by Duffin, Kemmer and others in connection with the mathematical theory of meson fields.[3] For this reason we shall call these systems *meson triple systems.* As we shall see, these systems are connected with the Lie algebras $\mathfrak{S}_{n+1}$ of $(n+1) \times (n+1)$ skew symmetric matrices. Systems that seem to be related in the same way with the Lie algebras of matrices of trace 0 are given in the next example.

2) Let $\mathfrak{K}$ be the space spanned by the matrices $x_i = e_{i,n+1}$, $y_i = e_{n+1,i}$. Here we have

$$\begin{aligned} (x_i y_j x_k) &= \delta_{jk} x_i + \delta_{ji} x_k, \\ (y_i x_j y_k) &= \delta_{jk} y_i + \delta_{ji} y_k \end{aligned} \tag{9}$$

with all other triples $(abc) = 0$. Thus $\mathfrak{K}$ is a J. t. s. For the associated L. t. s. we have

$$\begin{aligned} [[x_i, y_j], x_k] &= \delta_{jk} x_i + \delta_{ji} x_k \\ [[y_i, x_j], y_k] &= \delta_{jk} y_i + \delta_{ji} y_k. \end{aligned} \tag{10}$$

We consider next Lie triple systems that can be associated with arbitrary Jordan algebras. An algebra has been called a *Jordan algebra* by Albert[4] if it is commutative and its multiplication satisfies

$$(a^2 b) a = a^2 (ba). \tag{11}$$

Any special Jordan algebra is a Jordan algebra if the composition is taken to be $\{a, b\}$. On the other hand, there exist Jordan algebras that cannot be obtained as special Jordan algebras. Let R_a denote the right multiplication $x \to xa$ in a Jordan algebra $\mathfrak{J}$ and let L_a denote the left multiplication $x \to ax$. Then since $\mathfrak{J}$ is commutative $R_a = L_a$ and by (11) $R_{a^2} R_a = R_a R_{a^2}$.

---

[3] See the bibliography.

[4] Albert [2].

Also, as has been shown by Albert,[5] the latter relation can be linearized to give

$$R_{(ab)c} = R_a R_{bc} + R_b R_{ac} + R_c R_{ab} - (R_a R_c R_b + R_b R_c R_a) \tag{12}$$

if the characteristic is $\neq 2, 3$. This implies

$$R_{(ab)c-a(bc)} = [[R_c, R_a], R_b]. \tag{13}$$

Thus if $R(\mathfrak{J})$ is the set of right multiplications operating in $\mathfrak{J}$ then $R(\mathfrak{J})$ is a subspace of the algebra of linear transformations in the vector space $\mathfrak{J}$ closed under the Lie triple product $[[A, B], C]$. Hence $R(\mathfrak{J})$ is a Lie triple system. We shall call this system *the multiplication* L. t. s of $\mathfrak{J}$. ✓

If $\mathfrak{J}$ is a special Jordan algebra with an identity, the mapping $a \to R_a$ is $1 - 1$ and linear of $\mathfrak{J}$ into $R(\mathfrak{J})$. Since

$$\{\{a, b\}, c\} - \{a, \{b, c\}\} = [[c, a], b],$$

(13) implies that

$$R_{[[c,a],b]} = [[R_c, R_a], R_b].$$

Hence $a \to R_a$ is an isomorphism of the associator L. t. s. of $\mathfrak{J}$ onto the multiplication L. t. s.

3. Identities. We shall not attempt to obtain an axiomatic development of Lie or Jordan triple systems in this paper. However, we shall derive in this section a number of identities that may be useful for such a development. We consider first relations for $[[a, b], c]$ and these will be clearer if we denote $[[a, b], c]$ by $[abc]$.

Skew symmetry of $[a, b]$ and the Jacobi identity imply

$$[abc] = -[bac] \tag{14}$$

$$[abc] + [bca] + [cab] = 0. \tag{15}$$ ✓

The Jacobi identity

$$[[[a, b], [c, d]]; e] + [[[c, d], e]; [a, b]] + [[e, [a, b]]; [c, d]] = 0$$

becomes by using

$$[[a, b], [c, d]] = [[[a, b], c], d] - [[[a, b], d], c], \tag{16}$$

$$[[[[a, b], c], d], e] + [[[[b, a], d], c], e] + [[b, a]; [[c, d], e]] + [[d, c]; [[a, b], e]] = 0$$

---

[5] Albert [2], p. 299. Cf. also Jordan, Wigner and von Neumann [9]. An equation equivalent to (13) below is also given in Albert's paper [2], p. 250.

or

$$(17) \qquad [[abc]de] + [[bad]ce] + [ba[cde]] + [cd[abe]] = 0.$$

Next we apply (16) and $[[a, b], [c, d]] = -[[c, d], [a, b]]$ to derive

$$(18) \qquad [[[a, b], c], d] + [[[b, a], d], c] + [[[d, c], b], a] + [[[c, d], a], b] = 0.$$

Taking commutators with e we obtain

$$(19) \qquad [[abc]de] + [[bad]ce] + [[dcb]ae] + [[cda]be] = 0.$$

Also it follows directly from (16) that

$$[[[a, b], [c, d]], [e, f]] = [[[abc]de], f] + [[[bac]df], e] + [[[bad]ce], f] - [[[bad]cf], e].$$

Hence

$$[[[[a, b], [c, d]], [e, f]], g] = [[[abc]de]fg] + [[[bac]df]eg] + [[[bad]ce]fg] + [[[abd]cf]eg].$$

By Jacobi's identity for $[[x, y], z]$ where $x = [a, b]$, $y = [c, d]$, $z = [e, f]$,

$$(20) \qquad [[[abc]de]fg] + [[[bac]df]eg] + [[[bad]ce]fg] + [[[abd]cf]eg] + Q + R = 0$$

where Q and R are obtained from the four displayed terms by cyclically permuting the pairs (a, b), (c, d), (e, f).

We establish next an identity which shows that if $\mathfrak{J}$ is a special Jordan algebra, then $[a, b]^2 \varepsilon \mathfrak{J}$ for every $a, b \varepsilon \mathfrak{J}$.[6] It is well known that the mapping $x \to [x, b]$ has the formal properties of differentiation. Hence

$$[a^2, b] = a[a, b] + [a, b]a$$

and

$$[[a^2, b], b] = 2[a, b]^2 + a[[a, b], b] + [[a, b], b]a$$

or

$$(21) \qquad [a^2bb] = 2[a, b]^2 + \{a, [abb]\}$$

since $\mathfrak{J}$ is a L. t. s. $[abb]$ and $[a^2bb] \varepsilon \mathfrak{J}$. Hence by (20), $[a, b]^2$ is in $\mathfrak{J}$.

We consider finally the function (abc) regarded as a function of a and c. We set $(abc) - \{a, c\}_b$. Then this product satisfies the conditions for the product in a Jordan algebra, that is

---

[6] This property is mentioned without proof in Jordan [8].

$$\{a, c\}_b = \{c, a\}_b \tag{22}$$

$$\{\{\{a, a\}_b, c\}_b, a\}_b = \{\{a, a\}_b, \{c, a\}_b\}_b. \tag{23}$$

These can be verified directly, or, we can obtain them from the following useful observation. If $\mathfrak{J}_3$ is a Jordan triple system and $b \varepsilon \mathfrak{J}_3$ then the set $\mathfrak{J}_3 b$ of multiples ab, $a \varepsilon \mathfrak{J}_3$ is a special Jordan algebra. For $(abc)b = \{ab, cb\} \varepsilon \mathfrak{J}_3 b$. Similarly $b\mathfrak{J}_3$ is a special Jordan algebra.

II. Universal Algebras.

4. Universal associative algebra of a system of equations.[7] Let X be a set of non-commuting indeterminates x_α and let $\mathfrak{F}[X]$ be the free polynomial algebra generated by the x's. A precise description of $\mathfrak{F}[X]$ is the following. Let $\mathfrak{X}$ be the free semigroup generated by the x_α. Thus the elements of $\mathfrak{X}$ are the monomials $x_{i_1}x_{i_2}\cdots x_{i_r}$, $x_{i_j} \varepsilon X$ and two of these are regarded as equal only if they look alike. Finally multiplication is defined by

$$(x_{i_1}x_{i_2}\cdots x_{i_r})(x_{j_1}x_{j_2}\cdots x_{j_s}) = x_{i_1}x_{i_2}\cdots x_{i_r}x_{j_1}x_{j_2}\cdots x_{j_s}.$$

We can now define $\mathfrak{F}[X]$ to be the semi-group algebra over our field Φ of the semi-group $\mathfrak{X}$.

If S is an arbitrary subset of $\mathfrak{F}[X]$ and $\mathfrak{B}$ is the two-sided (Φ-) ideal generated by S, then we shall call the algebra $\mathfrak{U} = \mathfrak{F}[X] - \mathfrak{B}$ *the universal associative algebra* of the set of equations

$$p(x_1, x_2, \cdots, x_r) = 0, \tag{24}$$

for $p \varepsilon S$. For the sake of simplicity we shall also denote the coset $x_i + \mathfrak{B}$ by x_i. Then any element of $\mathfrak{U}$ is a polynomial in these x_i and the relations (24) hold. Thus $\mathfrak{U}$ can also be regarded as the algebra obtained from the free algebra by imposing the above relations.

Suppose now that $\mathfrak{A}$ is any associative algebra over Φ that contains elements y_α such that $p(y_1, y_2, \cdots, y_r) = 0$ for every $p \varepsilon S$. We shall call the subalgebra generated by the y's *the enveloping algebra* $\mathfrak{E}$ of the y's. Evidently $\mathfrak{E}$ is the totality of polynomials in the y's with coefficients in Φ. It is easy to see that the correspondence $x_\alpha \rightarrow y_\alpha$ determines a unique homomorphism of $\mathfrak{U}$ onto $\mathfrak{E}$.

We apply these results now to Lie and Jordan triple systems. Let $\mathfrak{L}_3$ be a L. t. s. with basis $Y = \{y_\alpha\}$ and multiplication table

---

[7] An alternative discussion of universal algebras is given in Birkhoff and Whitman [3]. Cf. also Jacobson and Jacobson [7] for the case of special Jordan algebras.

$$[[y_\alpha, y_\beta], y_\gamma] = \Sigma \mu_{\alpha\beta\gamma\delta} y_\delta. \tag{25}$$

We form the universal algebra $\mathfrak{U}$ of the set of polynomials

$$[[x_\alpha, x_\beta], x_\gamma] - \Sigma \mu_{\alpha\beta\gamma\delta} x_\delta.$$

Then since the mapping $x_\alpha \rightarrow y_\alpha$ determines a homomorphism of $\mathfrak{U}$ on the enveloping algebra $\mathfrak{E}$ of the y's, it is clear that the x's are linearly independent in $\mathfrak{U}$. Since $[[x_\alpha, x_\beta], x_\gamma] = \Sigma \mu_{\alpha\beta\gamma\delta} x_\delta$ in $\mathfrak{U}$, it is also clear that the space spanned by the x's is a Lie triple system isomorphic to $\mathfrak{L}_3$. We shall now replace $\mathfrak{L}_3$ by this subspace of $\mathfrak{U}$ and we shall change our notation and denote the latter system as $\mathfrak{L}_3$.

Suppose now that $a \rightarrow \bar{a}$ is a homomorphism of $\mathfrak{L}_3$ on $\bar{\mathfrak{L}}_3$ a Lie triple system contained in the associative algebra $\bar{\mathfrak{A}}$. Then if $\bar{x}_\alpha$ is the correspondent of x_α, $[[\bar{x}_\alpha, \bar{x}_\beta], \bar{x}_\gamma] = \Sigma \mu_{\alpha\beta\gamma\delta} \bar{x}_\delta$. Hence, the correspondence $x_\alpha \rightarrow \bar{x}_\alpha$ can be extended to a homomorphism of $\mathfrak{U}$ onto the enveloping algebra $\bar{\mathfrak{E}}$ of the $\bar{x}_\alpha$. Evidently this homomorphism maps a into $\bar{a}$. Also it is clear that $\bar{\mathfrak{E}}$ is the enveloping algebra of $\bar{\mathfrak{L}}_3$. Thus, we see that the homomorphism of $\mathfrak{L}_3$ onto $\bar{\mathfrak{L}}_3$ can be extended in one and only one way to a homomorphism of $\mathfrak{U}$ onto the enveloping algebra $\bar{\mathfrak{E}}$ of $\bar{\mathfrak{L}}$.

The properties of $\mathfrak{U}$ that we have noted are characteristic. For let $\tilde{\mathfrak{U}}$ be any associative algebra that contains a Lie triple system $\tilde{\mathfrak{L}}_3$ isomorphic to $\mathfrak{L}_3$ under the correspondence $a \rightarrow \tilde{a}$. Suppose that 1) $\tilde{\mathfrak{U}}$ is the enveloping algebra of $\tilde{\mathfrak{L}}_3$ and, 2) any homomorphism $\tilde{a} \rightarrow \bar{a}$ of $\tilde{\mathfrak{L}}_3$ onto a L. t. s. $\bar{\mathfrak{L}}_3$ can be extended to a homomorphism of $\tilde{\mathfrak{U}}$ into the associative algebra that defines $\bar{\mathfrak{L}}_3$. Then the correspondences $a \rightarrow \tilde{a}$ and $\tilde{a} \rightarrow a$ can be extended to homomorphisms of $\mathfrak{U}$ onto $\tilde{\mathfrak{U}}$ and of $\tilde{\mathfrak{U}}$ onto $\mathfrak{U}$ respectively. It follows that both correspondences are isomorphisms. In this sense, $\mathfrak{U}$ is uniquely determined. In particular, we see that $\mathfrak{U}$ is independent of the choice of basis. We shall call $\mathfrak{U}$ *the universal associative algebra* of $\mathfrak{L}_3$.

In a similar fashion we can define the universal associative algebra of a Jordan triple system. In fact, it is clear that our discussion is applicable to other algebraic systems that can be defined by using polynomial functions in an associative algebra.

5. Universal Lie algebra of a Lie triple system. Let $\mathfrak{L}_3$ be a Lie triple system and let $\mathfrak{N}$ be the subspace of the given associative algebra that is spanned by the vectors $[a, b]$, a, b in $\mathfrak{L}_3$. If we refer to the relation (16) and recall that $[[a, b], c]$ and $[[a, b], d] \varepsilon \mathfrak{L}_3$ for a, b, c, d in $\mathfrak{L}_3$, then we see that $\mathfrak{N}$ is a Lie algebra. Also the subspace $\mathfrak{L} = \mathfrak{L}_3 + \mathfrak{N}$ is a Lie algebra since

$[a, b] \varepsilon \mathfrak{N}$ if a and $b \varepsilon \mathfrak{L}_3$ and $[[a, b], c] \varepsilon \mathfrak{L}_3$ if $[a, b] \varepsilon \mathfrak{N}$ and $c \varepsilon \mathfrak{L}_3$. Clearly $\mathfrak{L}$ is the smallest Lie subalgebra containing $\mathfrak{L}_3$. We shall call $\mathfrak{L}$ the *enveloping Lie algebra* of $\mathfrak{L}_3$. It is also clear that $\mathfrak{L}$ is finite dimensional if $\mathfrak{L}_3$ is finite dimensional. In fact, if $y_1, y_2, \cdots, y_n$ is a basis for $\mathfrak{L}_3$, then the elements y_i, $[y_i, y_j]$, $i < j$ are generators for $\mathfrak{L}$. Hence the dimensionality of $\mathfrak{L}$ does not exceed $n(n+1)/2$.

We shall call $\mathfrak{L}$ a *universal Lie algebra* for $\mathfrak{L}_3$ if any homomorphism $a \rightarrow \bar{a}$ of $\mathfrak{L}_3$ onto a L. t. s. $\bar{\mathfrak{L}}_3$ can be extended to a homomorphism of the enveloping Lie algebras. It is evident that if such an extension exists, then it is unique. Now it is easy to obtain a universal Lie algebra for any L. t. s. $\mathfrak{L}_3$. For this purpose we assume that $\mathfrak{L}_3$ is contained in its universal associative algebra $\mathfrak{U}$. This can be arranged by replacing $\mathfrak{L}_3$ if necessary by an isomorphic copy of it. It is clear from the properties of the universal associative algebra $\mathfrak{U}$ that the enveloping Lie algebra $\mathfrak{L}$ of $\mathfrak{L}_3$ in $\mathfrak{U}$ is a universal Lie algebra. Also, it is clear that if $a \rightarrow \bar{a}$ is an isomorphism of $\mathfrak{L}_3$ onto $\bar{\mathfrak{L}}_3$ and if the enveloping Lie algebras $\mathfrak{L}$ and $\bar{\mathfrak{L}}$ are universal, then the extended homomorphism is an isomorphism. In this sense the universal Lie algebra is uniquely determined.

Suppose now that $\mathfrak{L}_3$ is finite dimensional and that $\mathfrak{L}$ is its universal Lie algebra. Let $a \rightarrow \bar{a}$ be a homomorphism of $\mathfrak{L}_3$ on $\bar{\mathfrak{L}}_3$ and assume that the enveloping algebra $\bar{\mathfrak{L}}$ of $\bar{\mathfrak{L}}_3$ has the same dimensionality as $\mathfrak{L}$. Then the extended homomorphism of $\mathfrak{L}$ on $\bar{\mathfrak{L}}$ must be an isomorphism. Hence, in this case $\bar{\mathfrak{L}}$ is universal for $\bar{\mathfrak{L}}_3$. Thus, the universal Lie algebra for a finite dimensional L. t. s. can be characterized as an enveloping Lie algebra of maximum dimensionality.

If $\mathfrak{L}$ is the universal Lie algebra of $\mathfrak{L}_3$ the subalgebra $\mathfrak{N}$ of elements of the form $\Sigma[a, b]$, a, b in $\mathfrak{L}_3$ will be called the *complementary Lie algebra* of $\mathfrak{L}$. We proceed now to determine the universal Lie algebras and complementary Lie algebras for the examples listed in **2**.

For the meson Lie triple system we have the basis $g_i = e_{i,n+1} - e_{n+1,i}$. Hence $[g_i, g_j] = e_{ji} - e_{ij}$. The elements g_i, $[g_i, g_j]$ $i < j$ are linearly independent and constitute a basis for the Lie algebra $\mathfrak{S}_{n+1}$ of $(n+1) \times (n+1)$ skew symmetric matrices. It follows that $\mathfrak{S}_{n+1}$ is universal. The complementary Lie algebra has the basis $e_{ij} - e_{ji}$, $i, j = 1, 2, \cdots, n$; $i < j$. Hence it is isomorphic to the Lie algebra $\mathfrak{S}_n$.

We consider next the Lie triple system $\mathfrak{K}$ with basis $x_i = e_{i,n+1}$, $y_i = e_{n+1,i}$, $i = 1, 2, \cdots, n$. Here $[x_i, x_j] = 0 = [y_i, y_j]$ and $[x_i, y_j] = e_{ij} - \delta_{ij} e_{n+1,n+1}$. Thus the Lie algebra $\mathfrak{N}$ spanned by the vectors $[a, b]$, a, b in $\mathfrak{K}$ has the basis $[x_i, y_j]$ and the enveloping Lie algebra $\mathfrak{L}$ has dimensionality $2n + n^2$

$=(n+1)^2-1$. It is clear that $\mathfrak{L}$ coincides with the Lie algebra $\Phi'_{n+1,l}$ of matrices of trace 0. Now let $\bar{\mathfrak{K}}$ be any L. t. s. that is isomorphic to $\mathfrak{K}$ and let $\bar{x}_i$, $\bar{y}_j$ be a basis corresponding to the basis x_i, y_j. If we use the multiplication table (10) and the identity (18) for $a=x_i$, $b=\bar{y}_j$, $c=\bar{x}_k$, $d=\bar{x}_l$ we obtain

$$2\delta_{ij}[\bar{x}_k,\bar{x}_l]+\delta_{jk}[\bar{x}_i,\bar{x}_l]-\delta_{jl}[\bar{x}_i,\bar{x}_k]=0.$$

If we set $i=j=k$ and $l\neq j$ we obtain $3[\bar{x}_k,\bar{x}_l]=0$. Hence $[\bar{x}_k,\bar{x}_l]=0$. Similarly, $[\bar{y}_k,\bar{y}_l]=0$. It follows that the dimensionality of the enveloping Lie algebra $\bar{\mathfrak{L}}$ of $\bar{\mathfrak{K}}$ does not exceed $(n+1)^2-1$. Hence $\mathfrak{L}=\Phi'_{n+1,l}$ is universal for $\mathfrak{K}$. The complementary algebra has the basis $e_{ij}-\delta_{ij}e_{n+1,n+1}$, $i,j=1,2,\cdots,n$. Hence it is isomorphic to the Lie algebra Φ_{nl} of all $n\times n$ matrices.

6. Finiteness of dimensionality of the universal associative algebra for a Jordan triple system. Let $\mathfrak{J}_3$ be a Jordan triple system and suppose that $\mathfrak{J}_3$ is contained in its universal algebra $\mathfrak{U}$. We wish to show that if $\mathfrak{J}_3$ is finite dimensional, then $\mathfrak{U}$ is finite dimensional. Let $x_1, x_2,\cdots,x_n$ be a basis for $\mathfrak{J}_3$. We shall call a monomial in the x_i *irreducible* if it cannot be expressed as a linear combination of monomials of lower degree. Also we shall call two irreducible monomials *equivalent* if their difference can be expressed in terms of lower degree monomials.

LEMMA. *An irreducible monomial is of degree* $\leq \max(2,n)$ *in any of the* x_i.

Let u denote one of the x's and let the monomial have the form $\cdots u\cdots u\cdots u\cdots$. If a term x_ix_ju appears it can be replaced by $-ux_jx_i+$ linear terms. This process can be used to replace the monomial by an equivalent one in which any u appears two places to the left of its original position. Eventually we can get an equivalent monomial in which the u's are separated by at most one $x_i\neq u$. Three u's cannot occur consecutively since u^3 can be expressed as a linear combination of x's. Also no terms of the form ux_iu^2 or u^2x_iu can occur, since $ux_iu^2=-u^3x_i+$ terms of lower degree and $u^2x_iu=-x_iu^3+$ terms of lower degree. Thus if u^2 occurs, then no other u appears in the monomial. Finally we consider a product of the form $(ux_i)(ux_j)\cdots(ux_l)$. We have seen that $u\mathfrak{J}_3$ is a special Jordan algebra and it is clear that the vectors ux_i, $i=1,2,\cdots,n$ are generators. Thus $(ux_i)(ux_j)=-(ux_j)(ux_i)+$ terms of lower degree. Thus we can permute the factors ux_i. But also $(ux_i)^2$ is a sum of terms of lower degree. It follows that no more than n factors ux_i can occur. This concludes the proof.

Evidently this lemma implies that there are only a finite number of irreducible monomials. Hence we have the following.

THEOREM. *If $\mathfrak{J}_3$ is a Jordan triple system of finite dimensionality, then its universal associative algebra $\mathfrak{U}$ is finite dimensional.*

Of course, this implies that the enveloping associative algebra of any J. t. s. with a finite basis has a finite basis.

III. Representation Theory for Meson Triple Systems.

7. Svartholm's problem. With a view of applications to quantum physics, N. Svartholm[8] has proposed the problem of determining the structure of the universal associative algebra of the system of equations

$$[[x_i, x_j], x_k] = \delta_{ki}x_j - \delta_{kj}x_i \tag{26}$$

$$\phi(x_i) = 0 \tag{27}$$

where $i, j, k = 1, 2, \cdots, n$ and $\phi(\lambda)$ is a non-zero polynomial in $\Phi[\lambda]$. It was shown by Svartholm that if $\phi(\lambda)$ is of degree 2, then $\Phi(\lambda) = \lambda^2 + \frac{1}{4}$ and $\mathfrak{U}$ is an algebra of Clifford numbers.[9] If $\phi(\lambda)$ is of degree 3, then $\phi(\lambda) = \lambda^3 + \lambda$, so that $x_i{}^3 = -x_i$. By (26) if $i \neq j$

$$x_ix_j{}^2 - 2x_jx_ix_j + x_j{}^2x_i = -x_i. \tag{28}$$

Hence

$$x_jx_ix_j{}^2 - 2x_j{}^2x_ix_j = 0 \text{ and } -2x_jx_ix_j{}^2 + x_j{}^2x_ix_j = 0.$$

Hence $x_jx_ix_j{}^2 = 0 = x_j{}^2x_ix_j$. Multiplication by x_j, gives

$$x_jx_ix_j = 0, \qquad i \neq j. \tag{29}$$

Thus (28) becomes

$$x_ix_j{}^2 + x_j{}^2x_i = -x_i. \tag{30}$$

Next let i, j, k be $\neq$. Then by (26)

$$x_ix_jx_k - x_jx_ix_k - x_kx_ix_j + x_kx_jx_i = 0.$$

---

[8] Svartholm [14].

[9] Svartholm [14]. We assume in the sequel that our minimum polynomials have no constant term. Using this convention the minimum polynomial in this case should be regarded as $\lambda^3 + \frac{1}{4}\lambda$.

If we multiply on the left by x_i^2 and use (29) and $x_i^3 = -x_i$, we obtain

$$-x_i x_j x_k + x_i^2 x_k x_j x_i = 0. \tag{31}$$

Since

$$\begin{aligned} x_i^2 x_k x_j x_i &= -(x_k x_i^2 + x_k) x_j x_i = -x_k x_i^2 x_j x_i - x_k x_j x_i \\ &= x_k (x_j x_i^2 + x_j) x_i - x_k x_j x_i = -x_k x_j x_i, \end{aligned}$$

(31) implies that

$$x_i x_j x_k + x_k x_j x_i = 0, \qquad i, j, k \neq. \tag{32}$$

Thus we have the equations

$$(x_i x_j x_k) = -\delta_{ij} x_k - \delta_{kj} x_i \tag{33}$$

for all i, j, k. Conversely, these equations imply (26) and $x_i^3 = -x_i$. Thus we see that if $\phi(\lambda) = \lambda^3 + \lambda$ then $\mathfrak{U}$ can also be defined as the universal associative algebra of the meson J. t. s. The structure of this algebra has been determined by Svartholm using associative methods. In the remainder of this paper we shall determine the structure of $\mathfrak{U}$ for arbitrary $\phi(\lambda)$. In fact, we shall replace (27) by the weaker requirement

$$\phi(x_1) = 0. \tag{34}$$

Thus let $\mathfrak{U}$ be the universal algebra of (26) and (34). We assume that Φ is an algebraically closed field of characteristic 0. Let $x_1 = x$, $x_2 = y$, $[x_1, x_2] = z$. Then it follows from (26) that

$$[x, y] = z, \qquad [y, z] = x, \qquad [z, x] = y. \tag{35}$$

If we set $h_1 = \sqrt{-1}x$, $e = y + \sqrt{-1}z$, $f = y - \sqrt{-1}z$, then

$$[h_1, e] = e, \qquad [h_1, f] = -f, \qquad [f, e] = 2h_1. \tag{36}$$

Evidently if $\psi(\lambda) = \phi(-\sqrt{-1}\lambda)$ then $\psi(h_1) = 0$. Also by (36) if $g(\lambda)$ is any polynomial, then

$$[g(h_1), e] = e\Delta g(h_1), \qquad \Delta g(\lambda) = g(\lambda + 1) - g(\lambda).$$

Iteration of this equation gives

$$[\cdots[g(h_1), \overbrace{e], e], \cdots, e]}^{r} = e^r \Delta^r g(h_1).$$

Now take $g(\lambda) = \psi(\lambda)$ and let r be the degree of $\psi(\lambda)$. Then

$$0 = [\cdots[\psi(h_1), \overbrace{e], e], \cdots, e]}^{r} = e^r \Delta^r \psi(h_1).$$

Since $\Delta^r\psi(h_1) = r!$ this implies that $e^r = 0$. In a similar manner we can prove that $f^r = 0$. We now note the following.

LEMMA. *The universal algebra of any system of equations of the form* $[x_i, x_j] = \Sigma\gamma_{ijk}x_k$, $\phi_i(x_i) = 0$, $i, j, k = 1, 2, \cdots, n$ *has a finite basis.*

Proof. We define irreducible monomial and equivalence as in the preceding section. Let $\phi_i(\lambda)$ be of degree r. Then we assert that any irreducible monomial is of degree $< r_i$ in x_i. For if we use the relation $x_ix_j = x_jx_i + \Sigma\gamma_{ijk}x_k$ we can permute the x's and replace our monomial by the equivalent one $x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n}$. Also $x_i^{r_i}$ is a linear combination of the terms $x_i, x_i^2, \cdots, x_i^{r_i-1}$. Hence each $k_i < r_i$.

It follows from this lemma that the enveloping algebra of h_1, e, f is finite dimensional. Since this algebra includes x_2 and $[x_1, x_2]$, these elements satisfy algebraic equations. In a similar manner we can prove that every x_i and every $[x_i, x_j]$ is algebraic. Since the space determined by these elements is a Lie algebra the enveloping algebra $\mathfrak{U}$ of these elements is finite dimensional by the lemma. This proves the

THEOREM. *The universal algebra of the set of equations* (26) *and* (34) *is finite dimensional.*

We can now conclude that $\mathfrak{U}$ has a $1 - 1$ representation by finite matrices. If $x_i \rightarrow X_i$ in a representation of $\mathfrak{U}$ then $x_i \rightarrow X_i$, $[x_i, x_j] \rightarrow [X_i, X_j]$ determines a representation of the Lie algebra $\mathfrak{S}_{n+1}$ of skew symmetric matrices. For we know that the universal Lie algebra of $\mathfrak{M}$ is $\mathfrak{S}_{n+1}$. Also we have the relation $\phi(X_1) = 0$ in any representation of $\mathfrak{U}$. Conversely, if $x_i \rightarrow X_i$ in a representation of $\mathfrak{S}_{n+1}$ and $\phi(X_1) = 0$ then the X_i satisfy (26) and (34) and therefore they define a representation of $\mathfrak{U}$. Now it is well known that any representation of $\mathfrak{S}_{n+1}$ is completely reducible. Hence we see that there exists a completely reducible $1 - 1$ representation of $\mathfrak{U}$. This implies that $\mathfrak{U}$ is semi-simple and since Φ is algebraically closed, $\mathfrak{U}$ is a direct sum of complete matrix algebras. Also we know that the simple matrix components of $\mathfrak{U}$ are in $1 - 1$ correspondence with the different similarity classes of irreducible representations of $\mathfrak{U}$. Now it is clear that a representation of $\mathfrak{U}$ is irreducible if and only if the associated representation of $\mathfrak{S}_{n+1}$ is irreducible. Thus we have reduced the problem of determining the structure of $\mathfrak{U}$ to that of determining the irreducible representations of $\mathfrak{S}_{n+1}$ for which $\phi(X_1) = 0$. Explicitly we have the following relation: Let $R_1, R_2, \cdots, R_s$ be irreducible representations of $\mathfrak{S}_{n+1}$ satisfying $\phi(X_1) = 0$ and suppose that the R_i are not similar and that every irreducible representation for which $\phi(X_1) = 0$ is similar to one of the R_i. Then if N_i is the degree of R_i, $\mathfrak{U}$ is a direct sum of s complete matrix algebra $\Phi_{N_i}{}^{(i)}$. Hence the dimensionality of $\mathfrak{U}$ is $\sum_1^s N_i{}^2$.

In the remainder of this paper we shall apply the known theory of representations of $\mathfrak{S}_{n+1}$ to determine the degrees N_i.

8. Representation theory for $\mathfrak{S}_{n+1}$.[10] We distinguish odd and even dimensionality and accordingly we take $n+1=2\nu+1$ and $n=1=2\nu$. In place of $\mathfrak{S}_{n+1}$ we consider first the Lie algebra $\bar{\mathfrak{S}}_{n+1}$ of matrices A such that $S^{-1}A'S=-A$ where

$$S=\left[\begin{array}{c|cc} 1 & & \\ \hline & 0 & 1_\nu \\ & 1_\nu & 0 \end{array}\right], \qquad S=\begin{bmatrix} 0 & 1_\nu \\ 1_\nu & 0 \end{bmatrix} \tag{37}$$

in the two cases. The Lie algebra $\bar{\mathfrak{S}}_{n+1}$ is the set of matrices

$$\left[\begin{array}{c|cc} 0 & -b_2 & -b_1 \\ \hline b_1 & a_{11} & a_{12} \\ b_2 & a_{21} & a_{22} \end{array}\right], \qquad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \tag{38}$$

where

$$a'_{12}=-a_{12}, \qquad a'_{21}=-a_{21}, \qquad a_{22}=-a'_{11}. \tag{39}$$

It follows that the subset of matrices

$$h=\begin{bmatrix} 0 & & & & & & \\ & \lambda_1 & & & & & \\ & & \ddots & & & & \\ & & & \lambda_\nu & & & \\ & & & & -\lambda_1 & & \\ & & & & & \ddots & \\ & & & & & & -\lambda_\nu \end{bmatrix}, \qquad h=\begin{bmatrix} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_\nu & & & \\ & & & -\lambda_1 & & \\ & & & & \ddots & \\ & & & & & -\lambda_\nu \end{bmatrix} \tag{40}$$

is a maximal commutative subalgebra $\mathfrak{H}$. Also there exist elements $e_\alpha \,\varepsilon\, \bar{\mathfrak{S}}_{n+1}$ such that

$$\begin{aligned} [h, e_\alpha] &= \alpha e_\alpha, \\ [e_\alpha, e_{-\alpha}] &= h_\alpha \,\varepsilon\, \mathfrak{H} \end{aligned} \tag{41}$$

---

[10] The results stated in this section are due to E. Cartan [5] and H. Weyl [15]. All of them can be found in Weyl's second paper, pp. 342-353.

where the *roots* α are

$$\alpha = \pm \lambda_i \pm \lambda_j,\ i < j \text{ and } \alpha = \pm \lambda_i \qquad \text{for } n+1 = 2\nu + 1 \tag{42}$$
$$\alpha = \pm \lambda_i \pm \lambda_j,\ i < j \qquad \text{for } n+1 = 2\nu.$$

If we have a representation of $\mathfrak{S}_{n+1}$ we can replace it by a similar one in which h is represented by

$$H = \begin{pmatrix} \Lambda_1 & & & \\ & \Lambda_2 & & \\ & & \ddots & \\ & & & \Lambda_N \end{pmatrix} \tag{43}$$

where the Λ_i are linear forms in the λ_i. These forms are called the *weights* of the representation. If $\Lambda = \Sigma m_i \lambda_i$ is a linear form in the λ_i, the value of Λ for the λ_i that give h_a is denoted as Λ_a. Then it is easy to see that $\Lambda_a = \pm m_i \pm m_j$ if $\alpha = \pm \lambda_i \pm \lambda_j$ and $\Lambda_a = \pm 2m_i$ if $\alpha = \pm \lambda_i$. In particular, $\alpha_a = 2$ in either case. It is known that if Λ is a weight, then $2\Lambda_a/\alpha_a$ is an integer and $\Lambda - (2\Lambda_a/\alpha_a)\alpha$ is also a weight. It follows that the coefficients of any weight are either all integers or all halves of odd integers; and if Λ is a weight, then any linear form that can be obtained from Λ by an arbitrary permutation of the λ_i, or by replacing any λ_i by $-\lambda_i$, $n+1 = 2\nu - 1$, any even number of the λ_i by $-\lambda_i$, $n+1 = 2\nu$, is a weight.

One orders the weights lexicographically by stipulating that $\Lambda = \Sigma m_i \lambda_i > 0$ if the first non-zero m_i is > 0. Then it is known that an irreducible representation is determined up to a similarity by its highest weight. For this weight we have

$$m_1 \geq m_2 \geq \cdots \geq m_\nu \geq 0, \qquad n+1 = 2\nu + 1 \tag{44}$$
$$m_1 \geq m_2 \geq \cdots \geq m_{\nu-1} \geq | m_\nu |, \qquad n+1 = 2\nu.$$

Conversely, if $\Lambda = \Sigma m_i \lambda_i$ satisfies (44), then Λ is the highest weight of an irreducible representation. Hence there is a $1-1$ correspondence between the similarity classes of irreducible representations and the linear forms satisfying (44). The degree of the irreducible representation with highest weight $\Sigma m_i \lambda_i$ is

$$\begin{cases} N(m_1, m_2, \cdots, m_\nu) = \dfrac{P(l_1, l_2, \cdots, l_\nu)}{P(\nu - \frac{1}{2}, \cdots, 3/2, \frac{1}{2})} \\ P(u_1, u_2, \cdots, u_\nu) = \Pi_{u_i} \prod\limits_{i<j} (u_i^2 - u_j^2) \\ l_i = m_i + \nu - (2i - 1/2), \qquad n+1 = 2\nu + 1 \end{cases} \tag{45}$$

$$(46)\quad \begin{cases} N(m_1, m_2, \cdots, m_\nu) = \dfrac{P'(l_1, l_2, \cdots, l_\nu)}{P'(\nu - 1, \cdots, 1, 0)} \\ P'(u_1, u_2, \cdots, u_\nu) = \prod\limits_{i<j} (u_i{}^2 - u_j{}^2) \\ l_i = m_i + \nu - i, \qquad\qquad n + 1 = 2\nu. \end{cases}$$

We recall also that all the weights of an irreducible representation are obtained from a particular one by adding roots. Hence, if one of the weights has integer coefficients, all have integer coefficients. Finally if Λ is the highest weight of a representation and $\alpha > 0$ then

$$\Lambda, \Lambda - \alpha, \Lambda - 2\alpha, \cdots, \Lambda - 2(\Lambda_a/\alpha_a)\alpha$$

are weights of the representation.

We consider now the special case of $\mathfrak{S}_3$. Here we have the roots λ_1, $-\lambda_1$ and a basis h_1, $e = e_{\lambda_1}$, $f = e_{-\lambda_1}$ so that (36) holds. Let $h = \lambda_1 h_1$ and $h \to H$ in an irreducible representation with maximum weight $m_1\lambda_1$. Then $m_1 \geq 0$ and $2\Lambda_a/\alpha_a = 2m_1$. Hence

$$m_1\lambda_1, (m_1 - 1)\lambda, (m_1 - 2)\lambda_1, \cdots, - m_1\lambda_1$$

are weights of the representation. It is easy to see from the results quoted that these are the only weights that occur. The number m_1 is either an integer or half of an odd integer. The degree of the irreducible representation is the odd integer $2m_1 - 1$ in the former case and the even integer $2m_1 - 1$ in the latter. Correspondingly the minimum polynomial of the matrix H_1 representing h_1 is

$$(47)\quad \begin{aligned} &(\lambda^2 - m_1{}^2)(\lambda^2 - (m_1 - 1)^2) \cdots (\lambda^2 - 1)\lambda, \text{ if } m_1 \text{ is an integer} \\ &(\lambda^2 - m_1{}^2)(\lambda^2 - (m_1 - 1)^2) \cdots (\lambda^2 - \tfrac{1}{4}), \text{ if } m_1 \text{ is half odd.} \end{aligned}$$

If we have a second polynomial of the form (47a) determined by the integer $m'_1 \leq m_1$ then it is a factor of (47a). A similar remark holds for (47b). It follows that the minimum polynomial of H_1 for any representation either has the form (47a), (47b) or is a product of a polynomial of the first type by one of the second. Thus the minimum polynomial *without constant term*[11] is

$$(48)\quad \psi_1(\lambda) = \begin{cases} (\lambda^2 - m_1{}^2)(\lambda^2 - (m_1 - 1)^2) \cdots (\lambda^2 - 1)\lambda, & m_1 \text{ an integer} \\ (\lambda^2 - m_1{}^2)(\lambda^2 - (m_1 - 1)^2) \cdots (\lambda^2 - \tfrac{1}{4})\lambda, & m_1 \text{ half odd} \\ (\lambda^2 - m_1{}^2) \cdots (\lambda^2 - 1)\lambda(\lambda^2 - m_2)^2 \cdots (\lambda^2 - \tfrac{1}{4}), & \begin{matrix} m_1 \text{ integer} \\ m_2 \text{ half odd.} \end{matrix} \end{cases}$$

---

[11] From now on the minimum polynomials will be taken in this sense. We assume also that leading coefficients are always 1.

The three possibilities occur according as (a) the irreducible components all have odd degrees and $2m_1+1$ is the maximum one, (b) the irreducible components all have even degrees and $2m_1+1$ is the maximum one, (c) both odd and even degrees occur for the irreducible components and $2m_1+1$ is the largest odd and $2m_2+1$ the largest even one. This result gives a simple proof of the following.

LEMMA. *If X, Y and Z are three matrices such that $[X,Y]=Z$, $[Y,Z]=X$, $[Z,X]=Y$ then the minimum polynomials without constant term of X, Y and Z are all equal and have the form $\phi_i(\lambda)=\gamma\psi_1(\sqrt{-1}\,\lambda)$, γ a constant.*

Proof. As in (35) and (36) we set $H_1=\sqrt{-1}\,X$, $E=Y+\sqrt{-1}\,Z$, $F=Y-\sqrt{-1}\,Z$. Then $[H_1,E]=E$, $[H_1,F]=-F$, $[F,E]=2H_1$. Hence $h_1\rightarrow H_1$, $e\rightarrow E$, $f\rightarrow F$ defines a representation of $\mathfrak{S}_3$ and the minimum polynomial of H_1 is given by (48) and is determined by the degrees of the irreducible constituents of the representation. In the same way, we can set $H_1=\sqrt{-1}\,Y$ and $\sqrt{-1}\,Z$ in turn. This shows that $\sqrt{-1}\,X$, $\sqrt{-1}\,Y$, $\sqrt{-1}\,Z$ have the same minimum polynomial $\psi_1(\lambda)$. It follows that X, Y, Z have the minimum polynomial $\phi_1(\lambda)=\gamma\psi(\sqrt{-1}\,\lambda)$.

9. The structure of $\mathfrak{U}$. We consider now an irreducible representation of $\mathfrak{S}_{n+1}$. As before we set $g_i=e_{i,n+1}-e_{n+1,i}$ $i=1,2,\cdots,n$, so that $[g_i,g_j]=e_{ji}-e_{ij}$. Hence if $x=g_i$, $y=g_j$ and $z=[g_i,g_j]$ then $[x,y]=z$, $[y,z]=x$, $[z,x]=y$. If $g_i\rightarrow G_i$ in our representation, then the above lemma shows that all the G_i and the $[G_i,G_j]$ have the same minimum polynomial $\phi_1(\lambda)=\gamma\psi_1(\sqrt{-1}\,\lambda)$ where $\psi_1(\lambda)$ is given by (48). We shall now show that the third possibility given in (48) is excluded for an irreducible representation of $\mathfrak{S}_{n+1}$. For this purpose we introduce the matrices

$$(49)\qquad M=\left(\begin{array}{c|cc}\sqrt{2} & & \\ \hline & 1_\nu & 1_\nu \\ & \sqrt{-1}\,1_\nu & -\sqrt{-1}\,1_\nu\end{array}\right),\qquad M=\left(\begin{array}{cc}1_\nu & 1_\nu \\ \sqrt{-1}\,1_\nu & -\sqrt{-1}\,1\end{array}\right.$$

Then

$$(50)\quad M'M=\left(\begin{array}{c|cc}2 & & \\ \hline & 0 & 2\cdot 1_\nu \\ & 2\cdot 1_\nu & 0\end{array}\right),\quad M'M=\left(\begin{array}{cc}0 & 2\cdot 1_\nu \\ 2\cdot 1_\nu & 0\end{array}\right).$$

It follows that the mapping $A\rightarrow MAM^{-1}$ is an isomorphism of $\tilde{\mathfrak{S}}_{n+1}$ on $\mathfrak{S}_{n+1}$. This isomorphism sends the matrix h given by (40) into

$$\left(\begin{array}{c|c|c}
0 & & \\ \hline
 & 0 & \begin{matrix}-\sqrt{-1}\,\lambda_1 & & \\ & \ddots & \\ & & -\sqrt{-1}\,\lambda_\nu\end{matrix} \\ \hline
 & \begin{matrix}\sqrt{-1}\,\lambda_1 & & \\ & \ddots & \\ & & \sqrt{-1}\,\lambda_\nu\end{matrix} & 0
\end{array}\right),$$

$$\left(\begin{array}{c|c}
0 & \begin{matrix}-\sqrt{-1}\,\lambda_1 & & \\ & \ddots & \\ & & -\sqrt{-1}\,\lambda_\nu\end{matrix} \\ \hline
\begin{matrix}\sqrt{-1}\,\lambda_\nu & & \\ & \ddots & \\ & & \sqrt{-1}\,\lambda_1\end{matrix} & 0
\end{array}\right).$$

Thus the matrix

$$\Sigma\lambda_i\sqrt{-1}\,(e_{\nu+i,i}-e_{i,\nu+i})$$

plays the role of h. Hence we can suppose that the matrix H corresponding to this matrix in our representation has the form (43) where the weights Λ satisfy the conditions given in the preceding section.

In particular we see that the coefficients of the weights are either all integers or all half-odd integers. If we set $\lambda_1=1, \lambda_2=\cdots=\lambda_\nu=0$, we obtain the matrix H_1 corresponding to $\sqrt{-1}(e_{\nu+1,1}-e_{1,\nu+1})$. Correspondingly the weight $\Sigma m_i\lambda_i$ becomes the characteristic root m_1 of H_1. Thus the characteristic roots of H_1 are either all integers or all half-odd. Hence we have proved that either (48a) or (48b) holds. Also we see that m_1 in the formula is the same as the m_1 in the highest weight $\Lambda=\Sigma m_i\lambda_i$.

We return now to the problem of determining the structure of the universal algebra $\mathfrak{U}$ of (26) and (34). As we have shown this problem can be solved by determining the irreducible representation of $\mathfrak{S}_{n+1}$ for which $\phi(X_1)=0$ where $x_1=g_1\to X_1$. Because of the form of the minimum polynomial for the matrix corresponding to g_1 in an irreducible representation, we can easily reduce the consideration of the $\mathfrak{U}$ determined by any $\phi(\lambda)$ to the two cases in which

$$(51)\quad \phi(\lambda) = \begin{cases} (\lambda^2 + m^2)(\lambda^2 + (m+1)^2) \cdots (\lambda^2 + 1)\lambda, & m \text{ an integer} \\ (\lambda^2 + m^2)(\lambda^2 + (m+1)^2) \cdots (\lambda^2 + \frac{1}{4})\lambda, & m \text{ half-odd.} \end{cases}$$

From now on we denote the algebra associated with this polynomial by $\mathfrak{U}_m$, $m = \frac{1}{2}, 1, 3/2, \cdots$. Now the minimum polynomial $\phi_1(\lambda)$ must be a factor of $\phi(\lambda)$. Hence we see that for the irreducible representations we seek $\phi_1(\lambda)$ has the same form as $\phi(\lambda)$ with m_1 in place of m and $m_1 \leq m$. Our results show that the maximum weight $\Lambda = \Sigma m_i \lambda_i$ satisfies (44) and

$$(52)\qquad m_1 \leq m, \qquad m_1 \equiv m \pmod 1.$$

The degree of the irreducible representation is given by (45) and (46). This proves the following

THEOREM. *Let $\mathfrak{U}_m$ be the universal associative algebra over an algebraically closed field of characteristic 0 of the system of equations* (26) *and* (34) *where $\phi(\lambda)$ is given by* (51). *Then $\mathfrak{U}_m$ is a direct sum of complete matrix algebras that are in $1 - 1$ correspondence with the sets of linear forms $\sum_{i=1}^{\nu} m_i \lambda_i$, $\nu = [n + 1/2]$, such that*

$$m \geq m_1 \geq m_2 \geq \cdots \geq m_\nu \geq 0, \qquad n + 1 = 2\nu + 1$$

$$m \geq m_1 \geq m_2 \geq \cdots \geq m_{\nu-1} \geq |\, m_\nu \,|, \qquad n + 1 = 2\nu$$

$$m_i \equiv m \pmod 1.$$

The dimensionality of the matrix component corresponding to the form $\Sigma m_i \lambda_i$ is $N(m_1, \cdots, m_\nu)^2$ where $N(m_1, m_2, \cdots, m_\nu)$ is given by (45) *and* (46).

Remarks. If m is an integer, the linear form 0 is excluded since in the corresponding representation of $\mathfrak{S}_{n+1}$ and of $\mathfrak{U}_m$ the representing elements are all 0. The dimensionality of $\mathfrak{U}_m$ is, of course, $\Sigma N(m_1, \cdots, m_\nu)^2$.

10. Meson Jordan triple systems.

We shall now see how our results specialize to give the known structure of the universal associative algebra of the meson J. t. s.[12] As we have seen in 7 this algebra is the universal algebra $\mathfrak{U}_1$. Hence the possible maximum weights for the irreducible representations are

---

[12] Svartholm [14] and D. E. Littlewood [13] employ associative methods to derive these results. Cf. also Kemmer [11].

$$(53)\quad \begin{array}{ll} \lambda_1, \lambda_1+\lambda_2, \cdots, \lambda_1+\lambda_2+\cdots+\lambda_\nu, & n+1=2\nu+1 \\ \lambda_1, \lambda_1+\lambda_2, \cdots, \lambda_1+\lambda_2+\cdots+\lambda_\nu, \lambda_1+\cdots+\lambda_{\nu-1}-\lambda_\nu, & n+1=2\nu. \end{array}$$

It is known that the associated irreducible representations are those whose representation spaces are the spaces of skew symmetric tensors of ranks $1, 2, \cdots, \nu$, if $n+1=2\nu+1$ and the spaces of skew symmetric of ranks $1, 2, \cdots, \nu-1$ plus the two irreducible spaces into which the skew symmetric tensor of rank ν splits if $n+1=2\nu$. For the sake of brevity we shall refer to these representations as the *irreducible tensor representations*. The number of simple components of $\mathfrak{U}_1$ is ν or $\nu+1$ and their degrees are

$$\binom{n+1}{1}, \binom{n+1}{2}, \cdots, \binom{n+1}{\nu}, \quad n+1=2\nu+1$$
$$\binom{n+1}{1}, \cdots, \binom{n+1}{\nu-1}, \tfrac{1}{2}\binom{n+1}{\nu}, \tfrac{1}{2}\binom{n+1}{\nu}, \quad n+1=2\nu.$$

Hence the dimensionality of $\mathfrak{U}_1$ is

$$\begin{aligned} q_n = {} & \binom{n+1}{1}^2+\binom{n+1}{2}^2+\cdots+\binom{n+1}{\nu}^2, \quad n+1=2\nu+1 \\ & \binom{n+1}{1}^2+\cdots+\binom{n+1}{\nu-1}^2+\tfrac{1}{2}\binom{n+1}{\nu}^2, \quad n+1=2\nu. \end{aligned}$$

Using the formula $\sum_{i=0}^{n+1}\binom{n+1}{i}^2=2\binom{2n+1}{n}$ we see that in either case

$$(55)\qquad q_n=\binom{2n+1}{n}-1.$$

11. A special 1-1 representation of $\mathfrak{U}_m$. Let $p_0, p_1, p_2, \cdots, p_n$ be generators of the Clifford algebra Γ satisfying the multiplication table

$$(56)\quad \begin{array}{ll} p_r^2=p_0, \quad p_0p_r=p_r=p_rp_0, & r=0,1,2,\cdots,n \\ p_ip_j=-p_jp_i, & i,j=1,2,\cdots,n.^{13} \end{array}$$

Then $p_0=1$ is the identity in Γ and we can verify that

$$(57)\quad \begin{array}{ll} [[p_i,p_j],p_k]=-2(\delta_{ki}p_j-\delta_{kj}p_i) & \\ p_i^3=p_i, & i,j,k=1,2,\cdots,n. \end{array}$$

Hence if we set $q_i=\frac{1}{2}\sqrt{-1}\,p_i$ then

$$(58)\qquad [[q_i,q_j],q_k]=\delta_{ki}q_j-\delta_{kj}q_i, \qquad q_i^3+\tfrac{1}{4}q_i=0.$$

---

[13] This algebra may be regarded as the universal associative algebra of a certain Jordan algebra. Cf. Jacobson and Jacobson [7].

Also it is clear that the enveloping algebra of the q_i is Γ itself, and as is well-known the dimensionality of Γ is 2^n.

On the other hand, let us consider now the universal algebra $\mathfrak{U}_{\frac{1}{2}}$ associated with the polynomial $\phi(\lambda) = \lambda^3 + \frac{1}{4}\lambda$. The irreducible representations of this algebra have maximum weights

$$\begin{array}{llll} (59) & \frac{1}{2}(\lambda_1 + \lambda_2 + \cdots + \lambda_\nu), & & n+1 = 2\nu + 1 \\ & \frac{1}{2}(\lambda_1 + \lambda_2 + \cdots + \lambda_\nu), & \frac{1}{2}(\lambda_1 + \cdots + \lambda_{\nu-1} - \lambda_\nu), & n+1 = 2\nu \end{array}$$

and are the well-known spinor representations of $\mathfrak{S}_{n+1}$.[14] The dimensionality of $\mathfrak{U}_{\frac{1}{2}}$ can be computed by the formulas to be $(2^\nu)^2 = 2^n$ if $n + 1 = 2\nu - 1$ and $2(2^{\nu-1})^2 = 2^n$ if $n + 1 = 2\nu$. Now by (58), the q_i of Γ satisfy the relations imposed on the generators x_i of $\mathfrak{U}_{\frac{1}{2}}$. Hence since Γ and $\mathfrak{U}_{\frac{1}{2}}$ have the same dimensionality we can identify x_i with q_i and regard Γ as the universal algebra $\mathfrak{U}_{\frac{1}{2}}$.

Our results now give the known structure of Γ: If $n + 1 = 2\nu + 1$, Γ is a complete matrix algebra of 2^ν rows and columns and if $n + 1 = 2\nu$ then Γ is a direct sum of two complete matrix algebras each of 2^ν rows and columns. In either case Γ has a $1 - 1$ representation S by matrices of 2^ν rows and columns.[15] We propose to show that this representation can be used to give a $1 - 1$ representation for $\mathfrak{U}_m$ for any m.

We recall that if R_1 and R_2 are two representations of a Lie algebra $\mathfrak{L}$ of respective degrees n_1 and n_2 then

$$(60) \qquad a \to a^{R_1 \times R_2} \equiv a^{R_1} \times 1_{n_2} + 1_{n_1} \times a^{R_2}$$

is a representation of $\mathfrak{L}$ of degree $n_1 n_2$. This representation is analogous to the direct product representation for groups and directly corresponds to the latter for Lie groups. It is evident that if R_1 is decomposed as

$$a^{R_1} = \begin{pmatrix} a^{R_{11}} & 0 \\ 0 & a^{R_{11}} \end{pmatrix}$$

then $R_1 \times R_2$ is decomposed as

$$\begin{pmatrix} a^{R_{11}} \times 1_{n_2} + 1_{n_{11}} \times a^{R_2} & 0 \\ 0 & a^{R_{12}} \times 1_{n_2} + 1_{n_{12}} \times a^{R_2} \end{pmatrix}$$

where n_{1i} is the degree of R_{1i}. Since $a \times b$ is similar to $b \times a$ a similar distributivity can be proved for the second factor. Hence, if R_1 is similar to a direct sum, $R_1 \sim R_{11} + R_{12} + \cdots + R_{1s}$ and $R_2 \sim R_{21} + R_{22} + \cdots + R_{2t}$ then $R_1 \times R_2 \sim \Sigma R_{1i} \times R_{2j}$.

---

[14] Cartan [5], p. 86; Brauer and Weyl [4].

[15] Explicit formulas for the representing matrices are given in Brauer and Weyl [4], p. 429 and p. 433.

Also it is clear that if h is represented by the diagonal matrix h^{R_1} with diagonal elements Λ_i and by the diagonal matrix h^{R_2} with diagonal elements M_j then $h^{R_1 \times R_2}$ is a diagonal matrix with diagonal elements $\Lambda_i + M_j$.

In particular we consider the representation S of $\mathfrak{S}_{n+1}$ (or of $\mathfrak{U}_{\frac{1}{2}}$). The weights of S are the 2^ν linear forms $\pm \frac{1}{2}\lambda_1 \pm \frac{1}{2}\lambda_2 \pm \cdots \pm \frac{1}{2}\lambda_\nu$. Hence the weights of the representation $S^2 = S \times S$ are the linear forms $0, \pm \lambda_1 \pm \lambda_2 \pm \cdots \pm \lambda_k$, $k \leq \nu$. The ones of these that can be maximal are $0, \lambda_1 + \lambda_2 + \cdots + \lambda_k$ for $n+1 = 2\nu + 1$ and $0, \lambda_1 + \lambda_2 + \cdots + \lambda_k$, $k \leq \nu$ and $\lambda_1 + \cdots + \lambda_{\nu-1} - \lambda_\nu$ for $n + 1 = 2\nu$. Hence, the possible irreducible constituents of S^2 are 0 and the irreducible tensor representations. It is known that all of these actually do occur. Hence the representation S^2 has the same irreducible constituents $\neq 0$ as the universal algebra $\mathfrak{U}_1$ of the meson J. t. s. It follows that S^2 gives a $1 - 1$ representation of $\mathfrak{U}_1$. We have, therefore, established the case $m = 1$ of the following

THEOREM. *If S denotes the $1 - 1$ representation of $\mathfrak{S}_{n+1}$ determined by the spinor representations, then the product S^{2m}*[16] *of $2m$ representations equivalent to S gives a $1 - 1$ representation of $\mathfrak{U}_m$, $m = \frac{1}{2}, 1, 3/2, \cdots$.*

Proof. Since the weights of S are $\pm \frac{1}{2}\lambda_1 \pm \frac{1}{2}\lambda_2 \pm \cdots \pm \frac{1}{2}\lambda_\nu$, the weights of the $2m$-fold product of S by itself have coefficients that are $\equiv m \pmod 1$ and are in absolute value $\leq m$. Hence the maximum weights of the irreducible constituents satisfy the conditions of the main theorem (9). It remains to show that every irreducible representation satisfying these conditions is a component of S^{2m}. We know that this holds for $m = \frac{1}{2}$ and $m = 1$ and we assume that it has already been proved for $m - 1$. Thus, any linear form satisfying the conditions of the main theorem for $m - 1$ in place of m occurs as the maximum weight of an irreducible component of S^{2m-2}. Also $\lambda_1 + \lambda_2 + \cdots + \lambda_k$, $0 \leq k \leq \nu$ and $\lambda_1 + \cdots + \lambda_{\nu-1} - \lambda_\nu$ if $n + 1 = 2\nu$ are maximum weights of the irreducible components of S^2. Now let $\Lambda = \Sigma m_i \lambda_i$ satisfy the conditions of the main theorem for the number m. Then $\Lambda = \Lambda' + (\lambda_1 + \cdots + \lambda_k)$ for some K, or $\Lambda = \Lambda' + (\lambda_1 + \cdots + \lambda_{\nu-1} - \lambda_\nu)$, $n + 1 = 2\nu$ where Λ' is maximum weight of an irreducible component in S^{2m-2}. Hence Λ is the maximum weight of a product of an irreducible component in S^{2m-2} and an irreducible component of S^2. It follows that Λ is a weight of an irreducible component of S^{2m}. This completes the proof.

---

[16] This type of representation of $\mathfrak{S}_{n+1}$ has been considered by Kramer, Belifante, and Lubanski [12] in generalizing spinors to obtain quantities called *undors*. A determination of the irreducible representations contained in $S \times S \times \cdots \times S$ for the case $n = 4$ has been given by Lubanski [14].

YALE UNIVERSITY.

BIBLIOGRAPHY.

1. A. A. Albert, "On Jordan algebras of linear transformations," *Transactions of the American Mathematical Society*, vol. 59 (1946), pp. 524-555.
2. ———, "A structure theory for Jordan algebras," *Annals of Mathematics*, vol. 48 (1947), pp. 546-567.
3. G. Birkhoff and P. Whitman, "Representation of Jordan on Lie algebras," to appear in the *Transactions of the American Mathematical Society.*
4. R. Brauer and H. Weyl, "Spinors in *n* dimensions," *American Journal of Mathematics*, vol. 57 (1935), pp. 425-449.
5. E. Cartan, "Les groupes projectifs qui ne laissent invariante aucune multiplicité plan," *Bulletin de la Société Mathematique de France*, vol. 41 (1913), pp. 53-96.
6. R. J. Duffin, "On the characteristic matrices of covariant systems," *Physical Review*, vol. 54 (1938), p. 1114.
7. N. Jacobson and F. D. Jacobson, "Structure and representation of semi-simple Jordan algebras," to appear in the *Transactions of the American Mathematical Society.*
8. P. Jordan, "Über die Multiplikation quanten-mechanischer Grössen," *Zeitschrift für Physik*, vol. 80 (1933), pp. 285-291.
9. P. Jordan, J. v. Neumann and E. Wigner, "On an algebraic generalization of the quantum mechanical formation," *Annals of Mathematics*, vol. 35 (1934), pp. 29-64.
10. N. Kemmer, "Particle aspect of meson theory," *Proceedings of the Royal Society*, vol. 173 (1939), pp. 91-116.
11. ———, "The algebra of meson matrices," *Proceedings of the Cambridge Philosophical Society*, vol. 39 (1943), pp. 189-196.
12. H. A. Kramers, F. J. Belinfante and J. K. Lubanski, "Über freie Teilchen mit nicht verschwindender Masse und beliebiger Spinquantinzahl," *Physica* 8, vol. 8 (1941), pp. 597-627.
13. D. E. Littlewood, "An equation of quantum mechanics," *Proceedings of the Cambridge Philosophical Society*, vol. 43 (1947), pp. 406-413.
14. J. K. Lubanski, "Sur la theorie des particules elementaire de spin quelconque," I and II, *Physica* 9 (1942), pp. 310-324 and 325-338.
15. N. Svartholm, "On the algebras of relativistic quantum theories," *Proceedings of the Royal Phisiographical Society of Lund*, vol. 12 (1942), pp. 94-108.
16. H. Weyl, "Theorie der Darstellung kontinuierlicher halbeinfacher Gruppen durch lineare Transformationen," I, II and III, *Mathematische Zeitschrift*, vols. 23-24 (1925-1926), pp. 271-304, 328-376, 377-395.

A CLASS OF INVERSE LIMIT GROUPS.*

By FRANKLIN HAIMO.

---

1. Introduction. Groups G to be considered here are to be additive abelian. The paper is concerned with certain inverse limit groups[1] of sets of quotient groups of G. The relation of these limit groups to related vector groups and to certain quotient groups of G will be studied. If G has at least one compact topology, the inverse limit groups may take on particularly simple forms. Recently, MacLane[2] has apparently used a special case of the type of limit group to be developed here.

Specifically, let $\mathfrak{H}$ be a directed system[1] of subgroups of a fixed group G, where $\mathfrak{H}$ is indexed by a directed set A, partially ordered by a relation $<$, in such a way that $\alpha \varepsilon A$, $\beta \varepsilon A$, $\alpha < \beta$ imply $H_\alpha \supset H_\beta$, and conversely. It follows that if H_α and H_β are in $\mathfrak{H}$, there exists $H_\gamma \varepsilon \mathfrak{H}$ with $H_\alpha \supset H_\gamma$, $H_\beta \supset H_\gamma$. There then exists a natural homomorphism for each ordered pair of elements of A which are in the relation $\alpha < \beta$, $\phi_\alpha{}^\beta$ of G/H_β onto G/H_α, with kernel isomorphic to H_α/H_β. This statement of the second isomorphism theorem, as applied to the groups of cosets of H_α and H_β in G, is, in part, $\phi_\alpha{}^\beta\colon g + H_\beta \rightarrow g + H_\alpha$, where $H_\alpha \supset H_\beta$, $g \varepsilon G$. If $\alpha < \beta < \gamma$, $\phi_\alpha{}^\beta \phi_\beta{}^\gamma = \phi_\alpha{}^\gamma$. If $\alpha < \beta$, we write $G/H_\alpha < G/H_\beta$. The system of quotient groups $\{G/H_\alpha\}$, $\alpha \varepsilon A$, under $<$ and with the ϕ's, forms an inverse limit group of a special sort. We denote it by invlim $(G; \mathfrak{H})$.

2. Canonical elements and compact groups. Let $\{g_\alpha + H_\alpha\}$ denote an element of invlim $(G; \mathfrak{H})$. Here, the g_α are in G and are subject to the relations $g_\gamma - g_\beta \varepsilon H_\beta$ whenever $H_\beta \supset H_\gamma$. Conversely, these relations imply the existence of an element of invlim $(G; \mathfrak{H})$. Suppose that there exists $g \varepsilon G$ such that $g - g_\alpha \varepsilon H_\alpha$ for every $\alpha \varepsilon A$. Then $\{g_\alpha + H_\alpha\} = \{g + H_\alpha\}$. Such an element of invlim $(G; \mathfrak{H})$ is called *canonical.* If all the elements

---

* Received April 16, 1948. Presented by title (of different wording) to the Society, April 16-17, 1948.

[1] Definitions and theorems concerning directed sets and inverse limit groups (in general) are to be found in S. Lefshetz, *Algebraic Topology,* New York, 1942. See especially pp. 4-5, 31-32, 54-56.

[2] S. MacLane, "The group of abelian group extensions," (an abstract), *Bulletin of the American Mathematical Society,* vol. 54 (1948), p. 53.

of invlim $(G;\mathfrak{H})$ are canonical, the inverse limit group, itself, is called canonical. Let the crosscut $\bigcap_{\alpha\varepsilon A} H_\alpha$ be denoted by H.

THEOREM I. *The set* $K(G;\mathfrak{H})$ *of canonical elements of* invlim $(G;\mathfrak{H})$ *is a subgroup (the canonical subgroup) of* invlim $(G;\mathfrak{H})$ *and is isomorphic to* G/H, *where* $H=\bigcap_{\alpha\varepsilon A} H_\alpha$.

Proof. The subgroup property is obvious. Suppose that $\{g+H_\alpha\}$ is a canonical element. Define a homomorphism O of $K(G;\mathfrak{H})$ onto G/H by $O(g+H_\alpha)=g+H$. If $\{g+H_\alpha\}=\{g'+H_\alpha\}$, $g-g'\,\varepsilon\, H_\alpha$ for every $\alpha\,\varepsilon\, A$ so that $g-g'\,\varepsilon\, H$, and $g+H=g'+H$. The converse statement also holds so that O is unique and has the kernel (0).

THEOREM II. *Let* O *be a homomorphism of* G *for which* kern $O\subset H=\bigcap_{\alpha\varepsilon A} H_\alpha$. *Then* invlim $(G;\mathfrak{H})$ *is canonical if, and only if,* invlim $(O(G);O(\mathfrak{H}))$ *is likewise canonical.* [Here, $O(\mathfrak{H})$ is a system of subgroups of $O(G)$, also directed by $\supset$ and indexed by A, and consists of the maps under induced O of the elements of $\mathfrak{H}$.]

Proof. Let $\{O(g_\gamma)+O(H_\gamma)\}$ be an element of invlim $(O(G);O(\mathfrak{H}))$. Then, if $H_\alpha\supset H_\beta$, $O(g_\alpha)-O(g_\beta)\,\varepsilon\, O(H_\alpha)$. Hence, there exist antecedents of $O(g_\alpha)$ and $O(g_\beta)$, g_α and g_β, respectively, such that $g_\alpha-g_\beta\,\varepsilon\, H_\alpha$, since the complete inverse image of $O(H_\alpha)$ is H_α. Thus, $\{g_\gamma+H_\gamma\}$ is an element of invlim $(G;\mathfrak{H})$. Suppose that invlim $(G;\mathfrak{H})$ is canonical. There exists $g\,\varepsilon\, G$ such that $g-g_\gamma\,\varepsilon\, H_\gamma$ for every $\gamma\,\varepsilon\, A$. It follows that $O(g)-O(g_\gamma)\,\varepsilon\, O(H_\gamma)$. Therefore, $\{O(g_\gamma)+O(H_\gamma)\}=\{O(g)+O(H_\gamma)\}$, so that invlim $(O(G);O(\mathfrak{H}))$ is canonical.

Conversely, suppose that invlim $(O(G);O(\mathfrak{H}))$ is canonical. Then, $\{g_\alpha+H_\alpha\}\,\varepsilon$ invlim $(G;\mathfrak{H})$ implies that $\{O(g_\alpha)+O(H_\alpha)\}$ is in invlim $(O(G);O(\mathfrak{H}))$. There exists $O(g)$ such that $O(g)-O(g_\alpha)\,\varepsilon\, O(H_\alpha)$ for every $\alpha\,\varepsilon\, A$. Hence, $g-g_\alpha\,\varepsilon\, H_\alpha$ for every α, so that $\{g_\alpha+H_\alpha\}=\{g+H_\alpha\}$; and invlim $(G;\mathfrak{H})$ is, consequently, canonical.

THEOREM III. *Let* G *possess at least one topology, making it a compact topological group in such a way that the members of a system of subgroups* $\mathfrak{H}$ *(directed by* $\supset$*), are closed. Then* invlim $(G;\mathfrak{H})$ *is canonical.*

Proof. Let an element of invlim $(G;\mathfrak{H})$ be $\{g_\alpha+H_\alpha\}$. Then the cosets $(g_\alpha+H_\alpha)$ are closed[3] as point-sets of G. Let $H_{\alpha_1}, H_{\alpha_2},\cdots,H_{\alpha_n}$ be any finite subset of $\mathfrak{H}$. There exists, (since $\mathfrak{H}$ is directed by $\supset$), $H_\beta\,\varepsilon\,\mathfrak{H}$,

---

[3] L. Pontrjagin, *Topological Groups*, Princeton, 1939. See p. 53.

$H_\beta \subset H_{\alpha_i}$, $i = 1, 2, \cdots, n$. Consequently, any finite subset $(g_{\alpha_i} + H_{\alpha_i})$, $i = 1, 2, \cdots, n$, of cosets has a non-void intersection. By compactness, this must also be true for the crosscut of all the cosets $(g_\alpha + H_\alpha)$, $\alpha \varepsilon A$. Hence there exists $g \varepsilon (g_\alpha + H_\alpha)$, for all $\alpha \varepsilon A$. Therefore, $\{g_\alpha + H_\alpha\} = \{g + H_\alpha\}$; and invlim $(G; \mathfrak{H})$ is, accordingly, canonical.

Let m and n be positive integers. If n is divisible by m, $mG \supset nG$. Using the notation [4] $G/sG = G_s$, we write $G_m < G_n$ if, and only if, $mG \supset nG$. It is well-known [5] that if G is compact then mG is closed for every positive integer m.

COROLLARY. *Let G' be the set of completely divisible elements of G. (That is, $G' = \bigcap_{m=1}^{\infty} mG$.) Let G have at least one topology making it a compact topological group. Then* invlim $(G; \{nG\})$ *is canonical and is isomorphic to G/G'. (Here $\mathfrak{H} = \{nG\}$ is the set of all nG, n a positive integer, directed by $\supset$.)*

Specializations of the groups used in the corollary have been studied by MacLane.[2]

Let [4] ${}_nG$ be the set of all $g \varepsilon G$ for which $ng = 0$. In a compact group the ${}_nG$ are closed. It follows that invlim $(G; \{{}_nG\})$ is canonical and isomorphic to G. However, this result holds even if G cannot be compact and is, indeed, a trivial consequence of the following theorem.

THEOREM IV. *Let $\mathfrak{H}'$ be a subset of $\mathfrak{H}$ which is cofinal therein. Let* invlim $(G; \mathfrak{H}')$ *be canonical. Then* invlim $(G; \mathfrak{H})$ *is canonical and is isomorphic to* invlim $(G; \mathfrak{H}')$.

Proof. This follows immediately from a well-known theorem on inverse limit groups in general.[1]

We say that the system $\mathfrak{H}$, indexed by A, contains a "last" element H_α under $\supset$ if $H_\beta \varepsilon \mathfrak{H}$ implies $H_\beta \supset H_\alpha$ for every $\beta \varepsilon A$.

COROLLARY. *If $\mathfrak{H}$ contains a last element under $\supset$, then* invlim $(G; \mathfrak{H})$ *is canonical.*

In the case of the ${}_nG$ above, ${}_1G = (0)$ is the last member of the set under $\supset$. Hence, always, G is isomorphic to invlim $(G; \{{}_nG\})$, and the latter is canonical. Another such example is furnished by the set $\mathfrak{H}$ of all subgroups of a group G which include a fixed subgroup H of G. Here, invlim $(G; \mathfrak{H})$ is canonical and is isomorphic to G/H.

---

[4] For some of the notation and for definitions see pp. 554-557 of P. Alexandroff and H. Hopf, *Topologie I*, Berlin, 1935.

Let the system $\mathfrak{H}$ be a refinement of the system $\mathfrak{H}'$. That is, $\mathfrak{H}'$ is to be a subset of $\mathfrak{H}$, and the partial ordering relation of the directed system $\mathfrak{H}'$ is to be extended to the partial ordering relation of the directed system $\mathfrak{H}$. Then there exists a homomorphism of invlim $(G; \mathfrak{H})$ into invlim $(G; \mathfrak{H}')$, a projection. The canonical subgroup of the former maps onto the canonical subgroup of the latter under this homomorphism.

3. A non-canonical inverse limit group.

We shall exhibit the existence of non-canonical inverse limit groups by a simple example. Let $p(1), p(2), \cdots$, be the enumeration of positive primes by increasing size. To each positive integer $s > 1$, there corresponds a sequence of non-negative integral exponents (almost all zero), $m(1,s), m(2,s), \cdots$, such that

$$s = \prod_{i=1}^{\infty} p(i)^{m(i,s)}.$$

Let $h(s)$, a non-negative integer, $0 \leqq h(s) < s$, be a solution of the simultaneous congruences

$$x \equiv i - 1 \mod p(i)^{m(i,s)}, \qquad i = 1, 2, \cdots.$$

$h(s)$ exists and is unique by the so-called Chinese remainder theorem.[6]

Let I be the group of integers. invlim $(I; \{sI\})$ has the element $\{h(s) + sI\}$. For, if t and u are positive integers greater than 1, then $t \mid u$ implies $m(i,t) \leqq m(i,u)$. It follows that

$$h(u) \equiv i - 1 \mod p(i)^{m(i,t)},$$
$$h(t) \equiv i - 1 \mod p(i)^{m(i,t)}.$$

Thus $h(u) - h(t) \equiv 0 \mod p(i)^{m(i,t)}$, $i = 1, 2, \cdots$, and this implies $h(u) - h(t) \in tI$.

On the other hand, $\{h(s) + sI\}$ is not canonical. Otherwise, there would exist $h \in I$ such that $h \equiv h(s) \equiv i - 1 \mod p(i)^{m(i,s)}$, $i = 1, 2, \cdots$; $s = 2, 3, \cdots$. But there exists $s \in I$, $s > 1$, for which $m(h,s) > 0$. For such s, $h \equiv h - 1 \mod p(h)$, and this is impossible.

A trivial consequence of this example is the well-known fact that I,

---

[5] S. Lefschetz, *op. cit.*, p. 68.

[6] See, for instance, p. 18 of C. C. MacDuffee, *An Introduction to Abstract Algebra*, New York, 1940.

the group of integers, can never be a compact topological group.[7] For, if it were, invlim $(I; \{nI\})$ would be canonical, by Theorem III, Corollary; and this is a contradiction.

The group invlim $(I; \{nI\})$ can be used to show that the result of Theorem II need not hold if the restriction on kern O is set aside. For, in the case of a proper homomorphic image $O(I)$ of I, invlim $(O(I); O(\{nI\}))$ is canonical while invlim $(I; \{nI\})$ is not. In fact, the kernel of no proper homomorphism of I can be included in $\bigcap_{n=1}^{\infty} nI$, since the latter is (0).

4. The division hull. The group invlim $(G; \mathfrak{H})$ is a subgroup of the cartesian product of all the G/H_α, $\alpha \varepsilon A$, the so-called strong direct sum or vector group of the G/H_α: $\sum_{\alpha \epsilon A} \oplus' G/H_\alpha$. Here, the symbol $\sum_{\alpha \epsilon A} 2 \oplus' B_\alpha$ stands for the cartesian product of the groups B_α with component-wise addition. It differs from the direct sum $\sum_{\alpha \epsilon A} \oplus B_\alpha$ in that elements with an infinite number of non-zero components are allowed. We should like to know something of the relation of the inverse limit group to the corresponding vector group. A few results are obtained here.

Following Alexandroff-Hopf,[4], H^0, the division hull of a subgroup H of G, is the set of all $g \varepsilon G$ such that there exists a non-zero integer $n(g)$ with $n(g) \cdot g \varepsilon H$. A subgroup H for which $H^0 = H$ is called a subgroup with division, (or H is said to have division in G).

THEOREM V. *If each $H_\alpha \varepsilon \mathfrak{H}$ is a subgroup with division in G, then* invlim $(G; \mathfrak{H})$ *has division in the corresponding vector group. Let $\mathfrak{H}$ have no last element under $\supset$; or let $\mathfrak{H}$ have a last element which is the intersection of all the other members of $\mathfrak{H}$. (In particular, the latter implies that $\mathfrak{H}$ has more than one member.) Then the converse of the above statement holds.*

Proof. Suppose that each H_α has division in G. Let $z = \{g_\alpha + H_\alpha\}$ be an element of $\sum \oplus' G/H_\alpha$, where the G/H_α are treated as groups of cosets. Suppose that $nz \varepsilon$ invlim $(G; \mathfrak{H})$. Then $\alpha < \beta$ implies $n(g_\alpha - g_\beta) \varepsilon H_\alpha$. It follows that $g_\alpha - g_\beta \varepsilon H_\alpha$. Therefore, $z \varepsilon$ invlim $(G; \mathfrak{H})$.

Conversely, let invlim $(G; \mathfrak{H})$ have division in the vector group. Suppose that $g \varepsilon G$, $ng \varepsilon H_\alpha$, where n is a non-zero integer. Construct z in the vector group such that the component of z in G/H_β is $0 + H_\beta$ if $H_\beta \not\supset H_\alpha$ and is $g + H_\beta$ if $H_\beta \supset H_\alpha$. nz is obviously the zero element of the vector group so that nz is in the inverse limit group. Therefore, $z \varepsilon \operatorname{invlim}(G; \mathfrak{H})$.

---

[7] For such questions, treated from a different point of view, see "Preservation of divisibility in quotient groups," *Duke Mathematical Journal*, vol. 15 (1948), pp. 347-356. Cf., A. Kurosh, "Primitive torsionfreie abelsche Gruppen vom endliche Range," *Annals of Mathematics* (2), vol. 38 (1937), pp. 175-203.

Case I. There exists $H_\gamma \neq H_\beta$, $H_\alpha \supset H_\gamma$. Then $0 + H_\gamma$ maps back onto $g + H_\alpha$. This implies $g \varepsilon H_\alpha$. *Case* II. $H_\alpha = \bigcap H_\beta$, $\beta \neq \alpha$. Then $ng \varepsilon H_\beta$, and by the method of proof for case I, $g \varepsilon H_\beta$ for every such $\beta \varepsilon A$. Hence, $g \varepsilon H_\alpha$.

COROLLARY. *Let* invlim $(G; \mathfrak{H})$ *be with division in* $\sum \oplus' G/H_\alpha$. *Then every non-last* (*under* $\supset$) H_α *is with division in* G.

The following example shows that Theorem V does not continue to hold if the restrictions on $\mathfrak{H}$ are dropped. Let $\mathfrak{H}$ have two members, I and $2I$, where I is the group of integers, and $G = I$. I has division in I, $2I$ does not, and $I \supset 2I$. The group invlim $(I; \mathfrak{H})$ is associated with the vector group $(0) \oplus I_2$, where $I_2 = I/2I$. Let n be a non-integer, $x \varepsilon (0)$, $y \varepsilon I_2$, and suppose that $n(x, y)$ is an element of invlim $(I; \mathfrak{H})$, where (x, y) is an element of the vector group. Let 1_2 be the only non-zero element of I_2. $x = 0$. But both $(0, 0)$ and $(0, 1_2)$, the only possibilities for (x, y), are in invlim $(I; \mathfrak{H})$ so that the latter has division in the vector group even though $2I$ does not have division in I.

THEOREM VI. *Let* $[\operatorname{invlim}(G; \mathfrak{H})]^0$ *be the division hull of* $\operatorname{invlim}(G; \mathfrak{H})$ *in the associated vector group. Let* $H_\alpha{}^0$ *be the division hull of* H_α *in* G. *Then there exists a homomorphism of* $[\operatorname{invlim}(G; \mathfrak{H})]^0$ *into* $\operatorname{invlim}(G; \{H_\alpha{}^0\})$, *the kernel of which consists of all those* $\{g_\alpha + H_\alpha\}$ *of* $[\operatorname{invlim}(G; \mathfrak{H})]^0$ *for which* $g_\alpha \varepsilon H_\alpha{}^0$ *for every* $\alpha \varepsilon A$.

Proof. If $\{g_\alpha + H_\alpha\} \varepsilon [\operatorname{invlim}(G; \mathfrak{H})]^0$, then there exists a positive integer m such that $m(g_\beta - g_\gamma) \varepsilon H_\beta$ whenever $\beta < \gamma$. Hence, $g_\beta - g_\gamma \varepsilon H_\beta{}^0$, and $\{g_\alpha + H_\alpha{}^0\}$ is in invlim $(G; \{H_\alpha{}^0\})$. The remaining details of the proof are trivial.

We note that the elements of $[\operatorname{invlim}(G; \{nG\})]^0$ are "almost" in the form of inverse limit elements. For, if $m\{g_n + nG\} \varepsilon \operatorname{invlim}(G; \{nG\})$, where $g_n \varepsilon G$, then $m(g_s - g_t) \varepsilon sG$ whenever $sG \supset tG$; and if $(m, s) = 1$, $g_s - g_t \varepsilon sG$.

Let H, a subgroup of G, have the property that G/H is a torsion group.[7] Then, to each $g \varepsilon G$, there exists a positive integer $n(g)$ such that $n(g) \cdot g \varepsilon H$, and conversely. This condition, also equivalent to the condition $H^0 = G$, might be described by saying that H is a *torsion generating* subgroup of G.

THEOREM VII. *Let* $\mathfrak{H}$ *have no last element under* $\supset$; *or, let the last element of* $\mathfrak{H}$ *be the intersection of all the other members of* $\mathfrak{H}$. *Then if* invlim $(G;\mathfrak{H})$ *be torsion generating in the corresponding vector group, each* H_α *is torsion generating in* G.

Proof. For $g \varepsilon G$ and for $\alpha \varepsilon A$, form an element z of the vector group in such a way that the component of z in G/H_α is $g + H_\alpha$ and in G/H_β is $0 + H_\beta$, for all $\beta \neq \alpha$. By hypothesis, there exists a positive integer n such that $nz \varepsilon \operatorname{invlim}(G;\mathfrak{H})$.

Case I. There exists $\beta \varepsilon A$ such that $H_\beta \neq H_\alpha$, $H_\alpha \supset H_\beta$. Then $0 + H_\beta$ maps back onto $ng + H_\alpha$, so that $ng \varepsilon H_\alpha$. *Case* II. $H_\alpha = \bigcap H_\beta$, $\beta \neq \alpha$. $ng + H_\alpha$ maps back onto $0 + H_\beta$ for all $\beta \neq \alpha$. It follows that $ng \varepsilon \bigcap H_\beta = H_\alpha$.

COROLLARY. *If* invlim $(G;\mathfrak{H})$ *is torsion generating in* $\sum \oplus' G/H_\alpha$, *then each non-last* H_α *is torsion generating in* G.

The converse of this theorem is not true, in general, even if we restrict $\mathfrak{H}$ as in the theorem. For example, in the group $\sum \oplus' I/nI$ choose an element z with n-th component $0 + nI$ for non-composite positive integers n and with n-th component $n/C + nI$ for composite positive integers n, where $C = C(n)$ is the product of all the first powers of the distinct positive prime divisors of n. If $z \varepsilon \operatorname{invlim}(I;\{nI\})$, $1 + 6I$ maps back onto $0 + 2I$, so that $1 \varepsilon 2I$, a contradiction. Thus, the inverse limit group is proper in the vector group. Consider all positive integral multiples mz of z, where $m > 1$. To each such m, there exists a positive, composite, odd integer n such that $(m, n) = 1$. Let p be the first prime smaller than any prime factor of n. Such a prime exists since n is odd. It follows that $(p - 1, n) = 1$ and that $(p, n) = 1$. Consider the components $n/C(n) + nI$ and $p^2n/C(p^2n) + p^2nI$ of z. Suppose that for some m as defined above $mz \varepsilon \operatorname{invlim}(I;\{sI\})$. Then $mn/C(n) - mp^2n/C(p^2n)$ should be in nI. But $C(p^2n) = pC(n)$, so that the above expression reduces to $mn(1 - p)/C$ which is not divisible by n. This proves that the inverse limit group is not torsion generating in the vector group, even though each sI is torsion generating in I. Note that the system of subgroups $\mathfrak{H} = \{sI\}$ has no last element under $\supset$.

An example to show that the conclusion of Theorem VII need not hold if $\mathfrak{H}$ has a last element is given by the group $I \oplus I$, the direct sum of two copies of the group I of integers. Let $\mathfrak{H}$ consist of two subgroups, $I \oplus I$ and $(0) \oplus I$. It is easy to show that the vector sum coincides with the inverse limit group so that the latter is torsion generating in the former. But $(0) \oplus I$ is not torsion generating in $I \oplus I$.

WASHINGTON UNIVERSITY,
ST. LOUIS, MO.

ALMOST PERIODIC INVARIANT VECTOR SETS IN A METRIC VECTOR SPACE.*

By Hermann Weyl.

---

1. The problem: basic notions and axioms. The situation underlying the theory of spherical harmonics deals with (complex-valued) functions $f(P)$ on the sphere Π under the influence of its rotations s. The rotations form a transitive group σ. The spherical harmonics of order l form a linear set of such functions (of dimensionality $2l+1$) that is invariant under all rotations. Here evidently the fact is used that functions $f(P)$ can be added and multiplied by numbers, in other words, that they form a vector space (of infinitely many dimensions). If the rotation s carries the point P into sP the transform $f' = sf$ of the function f is defined by $f'(sP) = f(P)$, or $sf(P) = f(s^{-1}P)$. It is obvious how to generalize this situation:

> I. Given a *group* σ and a *vector space* Σ. Any two vectors can be added and a vector can be multiplied by a number; for these operations the well-known axioms of vector geometry hold. Moreover there is associated with each element s of σ a linear transformation $f \to f' = sf$ of the vector space,
>
> $$s(f_1 + f_2) = (sf_1) + (sf_2), \qquad s(\alpha f) = \alpha \cdot sf,$$
>
> so that
>
> $$\mathsf{I} f = f, \qquad t(sf) = (ts)f$$
>
> (α being any number, f, f_1, f_2 any vectors, s, t any group elements and I denoting the unit element of σ).

Definition. h linearly independent vectors $g_1, \cdots, g_h$, or their linear combinations $\xi_1 \cdot g_1 + \cdots + \xi_h \cdot g_h$ by arbitrary numerical coefficients ξ_μ, form an invariant set Γ if $g \varepsilon \Gamma$ implies $sg \varepsilon \Gamma$ for every s in the group:

$$sg_\mu = \sum_\nu \omega_{\nu\mu}(s) \cdot g_\nu \qquad (\mu, \nu = 1, \cdots, h).$$

$g = \xi_1 g_1 + \cdots + \xi_h g_h$ is changed by s into $sg = \xi_1{}^s g_1 + \cdots + \xi_h{}^s g_h$ where

$$\xi_\mu{}^s = \sum_\nu \omega_{\mu\nu}(s)\xi_\nu.$$

---

* Received May 11, 1948.

$s \to \Omega(s) = \| \omega_{\mu\nu}(s) \|$ is a matric representation Ω of the group σ of degree h. Change of the basis of Γ changes Ω into an equivalent representation. We speak of an invariant set of degree h and order Ω. The set is called primitive if Ω is an irreducible representation.

The salient point in the theory of spherical harmonics is the fact that they form a complete orthogonal system. Clearly this fact depends on the possibility to form a scalar product (g, f) of two functions f, g (namely by integrating $\bar{g}(P) \cdot f(P)$ over the sphere by means of an area element $d\omega_P$ that is invariant under rotations). Hence we add to I the following assumptions:

> II. Σ is a metric vector space; i. e. with any two vectors f, g there is associated a number (g, f) as their scalar product, that is linear in the factor f, co-linear in the factor g:
>
> $$\begin{array}{ll|ll} (g, f_1 + f_2) = (g, f_1) + (g, f_2) & & & (g_1 + g_2, f) = (g_1, f) + (g_2, f) \\ (g, \alpha f) = \alpha \cdot (g, f) & & & (\alpha g, f) = \bar{\alpha} \cdot (g, f). \end{array}$$
>
> It is of Hermitian symmetry, $(f, g) = \overline{(g, f)}$, and $\| f \|^2 = (f, f)$ is positive except for $f = 0$. The linear transformations (s) in Σ, $f \to sf$, associated with the elements s of σ, are isometric, $(sg, sf) = (g, f)$.
>
> The equation $\| \alpha f \| = | \alpha | \cdot \| f \|$ and the inequalities
>
> $$|(f_2, f_1)| \leqq \| f_1 \| \cdot \| f_2 \|, \quad \| f_1 + f_2 \| \leq \| f_1 \| + \| f_2 \|$$

are an immediate consequence of these assumptions characteristic for a "unitary metric." More generally, for any numbers ξ_i and vectors f_i, g_i $(i = 1, \cdots, n)$ one has the relations

$$\| \sum_i \xi_i f_i \|^2 \leq \sum_i | \xi_i |^2 \cdot \sum_i \| f_i \|^2, \tag{1.1}$$

$$| \sum_i (g_i, f_i) |^2 \leq \sum_i \| g_i \|^2 \cdot \sum_i \| f_i \|^2, \tag{1.2}$$

which are proved in the same manner as the analogous Cauchy-Schwarz inequality for numbers.

We can now choose the basis g_μ $(\mu = 1, \cdots, h)$ of a given invariant set Γ as a unitary one, $(g_\mu, g_\nu) = \delta_{\mu\nu}$. Then $\Omega(s)$ is a unitary matrix and Ω a unitary representation. Every vector f may be split into a vector $\sum_\mu \alpha_\mu g_\mu$ $(\mu = 1, \cdots, h)$ lying in Γ and one f' that is perpendicular to Γ, by choosing α_μ as the Fourier coefficients of f, $\alpha_\mu = (g_\mu, f)$. Then

$$\| f \|^2 = \sum_\mu | \alpha_\mu |^2 + \| f' \|^2.$$

In the Parseval equation, the proof of which is our goal, only invariant sets of such functions will occur as can be prepared by linear combination from the transforms sf of f. But, of course, finite combinations alone will not do; we need some property of closure by which to pass from finite sums to their limits (integrals or mean values). But how shall we measure the degree to which a vector difference g approaches zero? If $\| g \|$ is used as this measure, then the continuous functions on the sphere will not form a closed vector space, one would have to include all Lebesgue-square-integrable functions. But it seems foolish to operate in such a wide field when all functions arising in the course of our construction are continuous. Therefore we introduce axiomatically the length $| g |$ of an arbitrary vector g as a quantity that does not necessarily coincide with the modulus $\| g \|$ and define closure in terms of this new notion.

> III. *Length and Closure.* Let there be associated with every vector f a non-negative number $|f|$, the length of f, which is zero for $f = 0$ and satisfies the conditions
>
> $$|f_1 + f_2| \leqq |f_1| + |f_2|, \qquad |\alpha f| = |\alpha| \cdot |f|,$$
> $$\|f\| \leqq |f|. \tag{1.3}$$
>
> The transform sf is supposed to have the same length as f, $|sf| = |f|$. Any sequence $f_1, f_2, \cdots$ of vectors is said to converge strongly if $|f_n - f_m| \to 0$ for $n, m \to \infty$. We require that any strongly convergent sequence f_n converges strongly toward a definite vector f, $|f_n - f| \to 0$ for $n \to \infty$.

It ought to be observed that, whether the space Σ is closed in this sense or not, it can always be extended into a closed space without violating any of the previous axioms. This is done by Cantor's classical procedure as follows: Any strongly convergent sequence f_n is said to define an ideal vector f^*; and two such sequences f_n, g_n are said to define the same f^* if $|f_n - g_n| \to 0$. Addition, multiplication by numbers, the definitions of (g, f), $|f|$ and of the transformation $f \to sf$ readily extend to these ideal vectors. In this way the axiom of closure can be *enforced* if it is not assumed.

We call a sequence f_n weakly convergent if $\| f_n - f_m \| \to 0$ for $n, m \to \infty$. The axioms about length are certainly fulfilled if we define $|f|$ as $\|f\|$. Strong-convergence-closure then becomes weak-convergence-closure.

Instead of Σ we may use as our field of operation any closed linear subspace Σ_f of Σ that is invariant with respect to all the transformations (s), $g \to sg$, and contains the given vector f. The "smallest" of these subspaces, $\Sigma_f{}^0$, consists of the strong limits of strongly convergent sequences, the members of which are finite linear combinations $\sum_i \xi_i \cdot s_i f$ of transforms sf of f. Our aim is to construct in $\Sigma_f{}^0$ a (finite or infinite) unitary sequence of vectors

$$g_n \quad (n = 1, 2, \cdot \cdot \cdot), \qquad (g_k, g_l) = \delta_{kl},$$

consisting of a string of invariant sets

$$g_1, \cdot \cdot \cdot, g_h \mid g_{h+1}, \cdot \cdot \cdot, g_{h+h'} \mid \cdot \cdot \cdot \quad \cdot \cdot \cdot$$

of such completeness that the Parseval equation

$$| \alpha_1 |^2 + | \alpha_2 |^2 + \cdot \cdot \cdot = \| f \|^2$$

holds for the Fourier coefficients $\alpha_n = (g_n, f)$ of f. This means that the partial sums of the Fourier series $\sum_n \alpha_n g_n$ converge *weakly* to f; for

$$\| f - \sum_{k=1}^{n} \alpha_k g_k \|^2 = \| f \|^2 - \sum_{k=1}^{n} | \alpha_k |^2.$$

A further axiom (Axiom IV in **4**) will be needed in order to ensure that f can also *strongly* be approximated with arbitrary accuracy by finite sums $\sum_k \beta_k g_k$.

No restriction shall be imposed on the group σ. However the proof of the fundamental theorem depends on an essential assumption concerning the vector f, an assumption to which for historical reasons the name "almost periodic" (abbreviated: a. p.) has been given. Using $| sf - tf |$ as a sort of distance between two elements s, t of the group, let us say that s lies in the f-circle of radius ϵ around a if $| sf - af | \leq \epsilon$. Any two elements s, t in this circle satisfy the relation $| sf - tf | \leq 2\epsilon$. The vector f is called *almost periodic* if the group is f-compact, i. e. if it may be covered by a finite number of f-circles of arbitrary small radius ϵ.

This definition may be replaced by the following criterion: f is a. p. if the group can be covered by a finite number of sets σ_i ("roof-tiles") such that $| sf - tf | \leq \epsilon$ for any two elements s, t in the same set σ_i. The σ_i are then said to form an (f, ϵ)-tiling. Indeed a finite number of f-circles of radius ϵ covering the whole group form an $(f, 2\epsilon)$-tiling. Vice versa, if the σ_i form an (f, ϵ)-tiling, we choose a "representative" a_i in each of the *non-empty* σ_i. The f-circles of radius ϵ around these representatives cover the group. Let f', f'' be two a. p. vectors, σ'_i $(i = 1, \cdot \cdot \cdot, n)$ an (f', ϵ)-tiling and σ''_k $(k = 1, \cdot \cdot \cdot, m)$ an (f'', ϵ)-tiling. Form the $n \cdot m$

intersections $\sigma_{ik} = \sigma'_i \cap \sigma''_k$. For any two elements s, t in the same σ_{ik} the inequalities

$$|sf' - tf'| \leq \epsilon, \qquad |sf'' - tf''| \leq \epsilon$$

hold simultaneously. In this sense the pair f', f'' is a. p. if each of the two members is. This fact carries over to any finite number of a. p. vectors $f_1, f_2, \cdots, f_l$. Hence not only the product of an a. p. vector by a number α, but also the sum of two a. p. vectors, and thus any linear combination $\sum_i \xi_i f_i$ of a. p. vectors is a. p. Moreover the strong limit of a strongly convergent sequence of a. p. vectors is a. p. Finally the transform af of an a. p. vector f by an element a of σ is a. p. Indeed let $a_1, \cdots, a_n$ be the centers of f-circles of radius ϵ covering the whole group, and set $a^*_i = a_i a^{-1}$. Then, for any given s, at least one of the inequalities

$$|saf - a^*_1 af| \leq \epsilon, \cdots, |saf - a^*_n af| \leq \epsilon$$

holds. We thus see that with f all the vectors of the space $\Sigma_f{}^0$ are a. p.

A numerical function $\phi(s)$ is f-continuous if for every $\delta > 0$ there exists a positive ϵ such that $|\phi(s) - \phi(t)| \leq \delta$ whenever $|sf - tf| \leq \epsilon$. For a vector function $g(s)$, whose values are vectors, the length $|g(s) - g(t)|$ has to replace the absolute value $|\phi(s) - \phi(t)|$ in this definition. The chief tool of the theory is the existence of a uniquely determined *mean value* $J = \int \phi(s)$ for every f-continuous function $\phi(s)$. A number of elements $a_1, \cdots, a_n$ and corresponding weights $\alpha_i \geq 0$ $(i = 1, \cdots, n)$ of sum 1 determine a (weighted) average

$$A = A\langle\phi\rangle = \sum_i^n \alpha_i \phi(a_i).$$

Let us say that this average oscillates by less than ϵ if

$$\Big|\sum_i \alpha_i \cdot \phi(aa_i) - A\Big| \leq \epsilon$$

for all elements a. It can be proved [Maak **6**; see also the Appendix of this paper] that there are such averages for any given positive ϵ, and that an average A_ϵ that oscillates by less than ϵ and an average B_δ that oscillates by less than δ differ by not more than $\epsilon + \delta$,

$$|A_\epsilon - B_\delta| \leq \epsilon + \delta. \tag{1.4}$$

Hence there exists a definite number J, the mean $\int \phi(s)$ of ϕ, such that $|A_\epsilon - J| \leq \epsilon$ for every average A_ϵ that oscillates by less than ϵ and every

$\epsilon > 0$. By its very definition the function $\phi(as)$ has the same mean J as $\phi(s)$ for any fixed elements a of σ. It can be shown that the same is true for $\phi(sa)$. All this carries over to an f-continuous vector function $g(s)$ of s.

The axioms I-III remain valid if σ is replaced by any subgroup σ^0 of σ. A vector f that is almost periodic with respect to σ is also a. p. with respect to σ^0. Indeed if a finite number of subsets σ_i of σ form an (f, ϵ)-tiling on σ then the sets $\sigma^0 \cap \sigma_i$ do so on σ^0.

Next we describe in precise terms that special interpretation of our axioms, to the simplest case of which the opening paragraph alluded.

> I_0. Given a group σ and a point field Π; moreover a realization of σ by transformations of Π. In other words, with every element s of σ there is associated a mapping $\{s\}: P \to P' = sP$ of Π upon itself such that $1P = P$, $t(sP) = (ts)P$. The group of these transformations is supposed to be transitive, i. e. given any two points P and Q there exists an element a of σ such that $Q = aP$. The transform sf of a numerical function $f = f(P)$ is defined by $sf(P) = f(s^{-1}P)$.

An element s is said to lie in the f-circle of radius ϵ around a if $| sf(P) - af(P) | \leq \epsilon$ identically in P. The function $f(P)$ is called almost periodic if the group can be covered by a finite number of f-circles of arbitrarily small radius ϵ. Choose a point P_0. This requirement obviously implies that $sf(P_0) = f(s^{-1}P_0)$ is a bounded numerical function of s, and hence, because of the transitivity of the group of transformations, $f(P)$ is a bounded function on Π. Looking upon the function $f(P)$ as a vector f in the functional space Σ of all a. p. functions on Π, we define $|f|$ as the lowest upper bound (l. u. b.) of $|f(P)|$, with the effect that our definition of almost-periodicity for the function $f(P)$ now coincides with that for the vector f as given above. Axioms I and III are satisfied in Σ with the (possible) exclusion of the relation (1. 3).

It remains to define the scalar product. For any given point P the numerical function $\psi(s) = sf(P)$ is evidently f-continuous, and we may form the average $\int \psi(s) = \int_s sf(P) = \int_s f(s^{-1}P)$. Since $\int \psi(as) = \int \psi(s)$, i. e. $\int_s f(s^{-1}Q) = \int_s f(s^{-1}P)$ for $Q = a^{-1}P$, this mean has the same value for P and for $Q = a^{-1}P$ and is thus a constant J. Write $J = \int_P f(Q)$ and call J the mean value of the a. p. function $f(P)$. The relation $\int \psi(sa) = \int \psi(s)$, on the other hand, shows that the transform af has the same mean as f. (It is only at this one place where the relation $\int \psi(sa) = \int \psi(s)$ comes into play.) The pair $f(P)$, $g(P)$ of two a. p. functions is a. p., and hence also the product $\bar{g}(P) \cdot f(P)$ of the conjugate of g by f. We now define the scalar product (g, f) as the integral $\int\limits_{P} \bar{g}(P) \cdot f(P)$ and then find that the axioms II and the relation (1. 3) are true in Σ. (In addition all the vectors of this space are almost periodic. But this is no serious restriction. For all our axioms carry over from a space Σ to the subspace of its almost periodic vectors.) We refer to this special interpretation as the "*interpretation by scalar functions on* Π."

It can be further specialized by identifying Π with σ and associating with the element a of σ the left translation $\{a\}$: $s \rightarrow as$ of $\Pi = \sigma$. The left translations constitute a transitive group of transformations on $\Pi = \sigma$ which is isomorphic to σ. This "*interpretation by scalar functions on* σ" is used in the construction of a complete set of inequivalent irreducible unitary representations of σ.

For compact Lie groups the theory of continuous representations was developed by F. Peter and the author [**8**] in 1926, and it was at once realized [**10**] that the method, I shall call it the integral equation method, carries over to the group of translations of a straight line and thus affords a natural approach to H. Bohr's theory of almost periodic functions [**2**]. It was J. von Neumann [**7**] who discovered that mean values can be defined for almost periodic functions on any group, and thus the construction of a complete set of a. p. representations was extended to arbitrary groups. In defining the mean value and deriving its essential properties we follow here a simplified procedure due to W. Maak [**6**]. The interpretation by scalar functions on Π (for compact Lie groups) was given by E. Cartan [**3**] and the author [**11**]. Results concerning unitary group representations in Hilbert space were obtained by S. Bochner and J. von Neumann [**1**] and by Anna Hurewitsch [**5**]. But the straightforward application of the integral equation method to the general situation staked out by our axioms seems to yield more precise and complete information.—No new methodological ideas will be developed in this paper; its purpose is to circumscribe the conditions under which a known method works.

2 gives the construction of the complete string of invariant sets in $\Sigma_f{}^0$ and thus derives the Parseval equation (weak approximation of f). The result is more fully evaluated in **3**, while **4** proves the central theorem concerning strong approximations, and adds a further interpretation in terms of "vector functions on Π." The Appendix (**5**) contains some remarks about Maak's procedure and the combinatorial lemma on which it is based.

2. The construction. From now on everything is relative to a given a. p. vector f so normalized that $\| f \|^2 \leq 1$.

The Cauchy-Schwarz inequality will be used by us in three different forms:

$$(C_1) \qquad \Big| \int_s \bar{\eta}(s)\xi(s) \Big|^2 \leq \int_s |\eta(s)|^2 \cdot \int_s |\xi(s)|^2,$$

$$(C_2) \qquad \Big\| \int_s \xi(s) \cdot g(s) \Big\|^2 \leq \int_s |\xi(s)|^2 \cdot \int_s \| g(s) \|^2,$$

$$(C_3) \qquad \Big| \int_s (g'(s), g(s)) \Big|^2 \leq \int_s \| g'(s) \|^2 \cdot \int_s \| g(s) \|^2.$$

$\xi(s)$, $\eta(s)$ are f-continuous numerical functions, $g(s)$, $g'(s)$ are f-continuous vector functions. (C_1) arises from Cauchy's inequality for numbers by passing from finite sums to the limit of integrals, (C_2) and (C_3) stand in the same relation to the inequalities (1. 1), (1. 2).

The given vector f defines the following linear mapping $\mathfrak{f}$ of the space Ξ of f-continuous functions $\xi(s)$ upon the vector space Σ,

$$\mathfrak{f}: \quad \xi(s) \rightarrow g = \int \xi(s) \cdot sf.$$

The image $g = \mathfrak{f}\xi$ lies in $\Sigma_f{}^0$. On the other hand, consider the linear mapping $\mathfrak{f}^*$ of the vector space Σ upon the space Ξ defined by

$$\mathfrak{f}^*: \quad g \rightarrow \xi(s) = (sf, g).$$

[$\xi(s)$ is not only f-continuous, but even satisfies a strong Lipschitz condition

$$(2.1) \qquad |\xi(s) - \xi(t)| \leq \text{const.}\ \| sf - tf \|.]$$

If $\int_s \bar{\eta}(s) \cdot \xi(s)$ is taken as the scalar product (η, ξ) in Ξ, the operator $\mathfrak{f}^*$ is the Hermitian conjugate of $\mathfrak{f}$. Indeed $\xi' = \mathfrak{f}^* g'$ gives $\bar{\xi}'(s) = (g', sf)$ and

$$(\xi', \xi) = \int_s \bar{\xi}'(s)\xi(s) = \int_s (g', sf) \cdot \xi(s) = (g', g)$$

where $g = \int_s \xi(s) \cdot sf$, hence $(\mathfrak{f}^* g', \xi) = (g', \mathfrak{f}\xi)$. Form the operator $\mathfrak{f}^*\mathfrak{f}$ in Ξ,

$$\xi(s) \rightarrow \int_t (sf, tf) \cdot \xi(t) = \int_t H(s, t) \cdot \xi(t),$$

with the Hermitian f-continuous kernel $H(s, t) = (sf, tf)$. This kernel is positive definite in the sense that

$$\int_s \int_t \bar{\xi}(s) \cdot H(s, t) \cdot \xi(t) \geq 0$$

for every function $\xi(s)$ in Ξ. Indeed the left side has the value

$$\Big\| \int_s \xi(s) \cdot sf \Big\|^2 = \| \mathfrak{f}\xi \|^2.$$

Because of its invariance (under left translations),

$$H(as, at) = H(s, t), \tag{2.2}$$

the kernel $H(s, t)$ actually depends on the one argument $t^{-1}s$ only. Note that

$$| H(s, t) | \leq \| sf \| \cdot \| tf \| = \| f \|^2 \leq 1.$$

The integral equation method consists in determining the (reciprocal) positive eigenvalues $\gamma, \gamma_1, \cdots$ of this kernel by E. Schmidt's procedure and thereby proving the equation

$$\operatorname{tr}(H) = \int_s H(s, s) = \gamma + \gamma_1 + \cdots.$$

The iterations of the kernel H are formed according to

$$H_1(s, t) = H(s, t), \qquad H_{n+1}(s, t) = \int_r H_n(s, r) H(r, t) \qquad (n = 1, 2, \cdots).$$

Set $\Gamma_n = \operatorname{tr}(H_n)$. The largest positive eigenvalue γ is constructed as the limit of the quotient Γ_{n+1}/Γ_n for $n \to \infty$, and the corresponding eigenkernel

$$E(s, t) = \phi_1(s) \cdot \bar{\phi}_1(t) + \cdots + \phi_h(s) \cdot \bar{\phi}_h(t) \tag{2.3}$$

as the limit of $H_n(s, t)/\gamma^n$ for $n \to \infty$. $\phi_1(s), \cdots, \phi_h(s)$ form a unitary basis for the eigenfunctions of $H(s, t)$ that belong to the eigenvalue γ. In this construction the invariance property (2.2) plays no rôle. However the latter is decisive for the fact that the h functions $\phi_\mu(s)$ and the corresponding vectors $g_\mu = \mathfrak{f}\phi_\mu$ constitute invariant sets.

In carrying out this program one has first to establish the inequalities

$$(2.4) \qquad \Gamma_{m+n} \leq \Gamma_m \cdot \Gamma_n, \qquad\qquad (2.5) \qquad \Gamma_n^2 \leq \Gamma_{n-l} \cdot \Gamma_{n+l}$$

$(m, n \geq 1,\ 0 \leq l < n)$, and this is the only point where the original exposition requires a slight new touch. Set

$$f_0(s) = sf, \qquad f_n(s) = \int_t H_n(s, t) \cdot tf \qquad (n = 1, 2, \cdots).$$

Then

$$\Gamma_{2n} = \int_s \int_t H_n(s, t) H_n(t, s) = \int_s \int_t | H_n(s, t) |^2 \geq 0 \quad (n = 1, 2, \cdots) \tag{2.6}$$

and

$$(2.7) \qquad \Gamma_{2n+1} = \int_s \| f_n(s) \|^2 \geq 0 \qquad (n = 0, 1, \cdots).$$

The second relation follows from the general equation

$$(2.8) \qquad (f_m(s) . f_n(t)) = H_{m+n+1}(s, t) \qquad (n, m \geq 0).$$

Hence all $\Gamma_n \geqq 0$ $(n = 1, 2, \cdots)$. In proving (2.4) we distinguish the three cases (i) m, n even; (ii) m even, n odd; (iii) m, n odd.

(i). $H_{m+n}(s, t) = \int_r H_m(s, r) H_n(r, t)$ yields by means of the inequality (C_1):

$$| H_{m+n}(s, t) |^2 \leq \int_r | H_m(s, r) |^2 \cdot \int_r | H_n(r, t) |^2$$

and hence by integrating over (s, t) and using the definition (2.6):

$$\Gamma_{2(m+n)} \leq \Gamma_{2m} \cdot \Gamma_{2n}.$$

(ii) $f_{m+n}(s) = \int_t H_m(s, t) f_n(t)$ gives by means of (C_2):

$$\| f_{m+n}(s) \|^2 \leq \int_t | H_m(s, t) |^2 \cdot \int_t \| f_n(t) \|^2,$$

and thus in view of (2.6), (2.7), after integration over s,

$$\Gamma_{2(m+n)+1} \leq \Gamma_{2m} \cdot \Gamma_{2n+1}.$$

(iii) From (2.8) there follows

$$| H_{m+n+1}(s, t) |^2 \leq \| f_m(s) \|^2 \cdot \| f_n(t) \|^2$$

and then by integration over s and t,

$$\Gamma_{2(m+n+1)} \leq \Gamma_{2m+1} \cdot \Gamma_{2n+1}.$$

In (2.5) we distinguish the two cases (i) $n \pm l$ even; (ii) $n \pm l$ odd. Set in the first case $n - l = 2u$, $n + l = 2v$, in the second $n - l = 2u + 1$, $n + l = 2v + 1$.

$$\text{(i)} \qquad \Gamma_n = \int_s H_{u+v}(s, s) = \int_s \int_t H_u(s, t) H_v(t, s),$$

hence application of (C_1) to integrals over the pair (s, t) gives the desired result

$$| \Gamma_n |^2 \leq \int_s \int_t | H_u(s, t) |^2 \cdot \int_s \int_t | H_v(t, s) |^2 = \Gamma_{2u} \cdot \Gamma_{2v}.$$

(ii) Apply (C_3) to

$$\Gamma_n = \int_s H_{u+v+1}(s,s) = \int_s (f_u(s), f_v(s))$$

with the result

$$|\Gamma_n|^2 \leq \int_s \|f_u(s)\|^2 \cdot \int_s \|f_v(s)\|^2 = \Gamma_{2u+1} \cdot \Gamma_{2v+1}.$$

Only the case $l = 1$ of (2.5) will be used,

$$\Gamma_n^2 \leq \Gamma_{n-1} \cdot \Gamma_{n+1} \qquad (n = 2, 3, \cdots). \tag{2.9}$$

If $f \neq 0$ then $\Gamma_1 = \|f\|^2 > 0$ and also $\Gamma_2 = \int_s\int_t |(sf, tf)|^2 > 0$ [for otherwise $(sf, tf) = 0$ for all s and t, in particular $(f, f) = 0$]. The inequality (2.9) then shows that one after the other of the traces $\Gamma_3, \Gamma_4, \cdots$ are likewise positive and not zero. More precisely we find that $q_n = \Gamma_{n+1}/\Gamma_n$ $(n = 1, 2, \cdots)$ is an increasing sequence, while (2.4) shows that

$$q_n^m \leq q_n q_{n+1} \cdots q_{n+m-1} = \Gamma_{n+m}/\Gamma_n \leq \Gamma_m, \tag{2.10}$$

or that the sequence q_n never grows beyond $\Gamma_m^{1/m}$, in particular not beyond $\Gamma_1 \leq 1$. Hence it converges to a positive number γ. Note in particular the inequality $q_1 \leq \gamma$ or

$$\Gamma_2 \leq \gamma\Gamma_1. \tag{2.11}$$

It is even easy to make an explicit estimate of the rapidity of convergence. Considering that q_n does not grow beyond $\Gamma_n^{1/m}$ and $\Gamma_m = \Gamma_1 q_1 \cdots q_{m-1}$ $\leq q_m^{m-1}$ one finds that q_n from $n = m$ on can not grow by more than

$$q_m^{(m-1)/m}(1 - q_m^{1/m}) \leq 1 - q_m^{1/m} \leq (q_m^{-1} - 1)/m \leq (q_1^{-1} - 1)/m.$$

Γ_n/γ^n decreases with increasing n since $q_n = \Gamma_{n+1}/\Gamma_n \leq \gamma$, hence tends to a limit h. But as (2.10) in the limit for $n \to \infty$ gives $\gamma^m \leq \Gamma_m$, the limit h of Γ_m/γ^m for $m \to \infty$ is by necessity greater than or equal to 1. (As h will turn out to be the multiplicity of the eigenvalue γ, an explicit estimate can not be expected for this second step!) Application of (C_1) to

$$\frac{H_{m+2}(s,t)}{\gamma^{m+2}} - \frac{H_{n+2}(s,t)}{\gamma^{n+2}} = \frac{1}{\gamma^2} \int\int_{r,r'} H(s,r) \left\{ \frac{H_m(r,r')}{\gamma^m} - \frac{H_n(r,r')}{\gamma^n} \right\} H(r',t)$$

yields

$$\left| \frac{H_{m+2}(s,t)}{\gamma^{m+2}} - \frac{H_{n+2}(s,t)}{\gamma^{n+2}} \right|^2 \leq \frac{1}{\gamma^4} \left\{ \frac{\Gamma_{2m}}{\gamma^{2m}} - \frac{\Gamma_{m+n}}{\gamma^{m+n}} + \frac{\Gamma_{2n}}{\gamma^{2n}} \right\}$$

and thus the uniform convergence of $H_n(s,t)/\gamma^n$ to an f-continuous limit $E(s,t)$. Invariance, (2.2), carries over from H to E, $E(as, at) = E(s,t)$. From the equations

$$\int_r H(s,r)E(r,t) = \int_r E(s,r)H(r,t) = \gamma \cdot E(s,t);$$

$$\int_r E(s,r)E(r,t); \qquad \bar{E}(s,t) = E(t,s); \qquad \operatorname{tr}(E) = h$$

there follows by a well-known elementary argument such a relation as (2.3) in which the $\phi_\mu(s)$ form a unitary-orthogonal system of f-continuous eigenfunctions of H,

$$\text{(2.12)} \qquad \int_t H(s,t)\phi_\mu(t) = \gamma \cdot \phi_\mu(s),$$

$$\int_s \bar{\phi}_\mu(s) \cdot \phi_\nu(s) = \delta_{\mu\nu} \qquad (\mu, \nu = 1, \cdots, h).$$

h turns out to be a positive integer. The one equation (2.12) may be split into two,

$$\text{(2.13)} \qquad \sqrt{\gamma} \cdot g_\mu = \int \phi_\mu(s) \cdot sf, \qquad \sqrt{\gamma} \cdot \phi_\mu(s) = (sf, g_\mu),$$

by using the first as definition of the vector g_μ. The g_μ are unitary-orthogonal,

$$(g_\mu, g_\nu) = \delta_{\mu\nu} \qquad (\mu, \nu = 1, \cdots, h).$$

$\mathfrak{f}\mathfrak{f}^* = \mathfrak{F}$ is a linear operator which carries the arbitrary vector g into

$$\mathfrak{F}g = \int_s (sf, g) \cdot sf.$$

This operator is invariant in the sense that

$$\text{(2.14)} \qquad \mathfrak{F}(ag) = a(\mathfrak{F}g) \qquad \text{for any group element } a.$$

Indeed

$$a(\mathfrak{F}g) = \int_s (sf, g) \cdot asf = \int_s (asf, ag) \cdot asf = \int_s (sf, ag) \cdot sf = \mathfrak{F}(ag).$$

The corresponding Hermitian form depending on two arbitrary vectors g, g' is

$$(g', \mathfrak{F}g) = \int_s (g', sf) \cdot (sf, g).$$

$(g', \mathfrak{F}g)$ is conjugate to $(g, \mathfrak{F}g')$. By combining the two equations (2.13) in inverse order one sees that the vectors g_μ obtained by our construction are eigenvectors of the operator $\mathfrak{F}$ for the eigenvalue γ,

$$\text{(2.15)} \qquad \mathfrak{F}g_\mu = \gamma \cdot g_\mu.$$

By starting with the equation

$$\phi_\mu(s) = \int_t E(s, t) \cdot \phi_\mu(t)$$

and utilizing the invariance property of E one finds that

$$\begin{aligned}\phi_\mu(a^{-1}s) &= \int_t E(a^{-1}s, t) \cdot \phi_\mu(t) = \int_t E(a^{-1}s, a^{-1}t) \cdot \phi_\mu(a^{-1}t) \\ &= \int_t E(s, t) \cdot \phi_\mu(a^{-1}t) = \sum_\nu \phi_\nu(s) \cdot \int_t \phi_\nu(t)\phi_\mu(a^{-1}t)\end{aligned}$$

is a linear combination $\sum_\nu \omega_{\nu\mu}(a) \cdot \phi_\nu(s)$ of the $\phi_\nu(s)$, and thus

$$\int_s \phi_\mu(s) \cdot asf = \int_s \phi_\mu(a^{-1}s) \cdot sf = \sum_\nu \omega_{\nu\mu}(a) \cdot \int_s \phi_\nu(s) \cdot sf$$

or $ag_\mu = \sum_\nu \omega_{\nu\mu}(a) \cdot g_\nu$. This equation

$$sg_\mu = \sum_\nu \omega_{\nu\mu}(s) \cdot g_\nu \tag{2.16}$$

shows that the g_μ form an invariant set $\{g_\mu\}$ belonging to the (necessarily unitary) representation Ω: $s \to \Omega(s) = \| \omega_{\mu\nu}(s) \|$. Let us say that h vectors g_μ $(\mu = 1, \cdots, h)$ transform according to Ω and call $(g_1, \cdots, g_h)$ briefly an Ω-row, if the equations (2.16) hold for every group element s. For any such row we infer from (2.16):

$$(sf, g_\mu) = (f, s^{-1}g_\mu) = \sum_\nu \omega_{\nu\mu}(s^{-1}) \cdot (f, g_\nu). \tag{2.17}$$

(f, g_ν) is the conjugate of the Fourier coefficient $\alpha_\nu = (g_\nu, f)$. Moreover the matrix $\| \omega_{\mu\nu}(s^{-1}) \|$ is reciprocal to $\| \omega_{\mu\nu}(s) \|$ and the latter is unitary, hence $\omega_{\nu\mu}(s^{-1}) = \bar{\omega}_{\mu\nu}(s)$, whereby (2.17) turns into

$$(sf, g_\mu) = \sum_\nu \bar{\omega}_{\mu\nu}(s) \cdot \alpha_\nu. \tag{2.18}$$

For two Ω-rows g, g' one gets

$$\sum_\mu (g'_\mu, sf) \cdot (sf, g_\mu) = \sum_\mu \alpha'_\mu \bar{\alpha}_\mu,$$

and hence by integration over s:

$$\sum_\mu (g'_\mu, \mathfrak{F}g_\mu) = \sum_\mu \alpha'_\mu \bar{\alpha}_\mu. \tag{2.19}$$

The special set $\{g_\mu\}$ constructed above satisfies the relation (2.15), thus $(g_\mu, \mathfrak{F}g_\mu) = \gamma$, and (2.19) yields the important equation

$$h\gamma = \sum_\mu \alpha_\mu \bar{\alpha}_\mu. \tag{2.20}$$

Let us say that the invariant set $\{g_\mu\}$ occurs in f with the (non-vanishing) *weight* $\gamma = h^{-1}\sum_\mu |\alpha_\mu|^2$.

When we subtract

$$e = \sum_{\mu=1}^{h} \alpha_\mu g_\mu$$

from f, the remainder $f' = f - e$ is orthogonal to the vectors g_μ. If it is still different from zero, we repeat for f' the process carried out before on f. One gets a new eigenvalue γ' and again a corresponding set of eigenfunctions $\phi'_{\mu'}(s)$ $(\mu' = 1, \cdots, h')$ and vectors $g'_{\mu'}$. The $\phi'_{\mu'}(s)$ turn out to be orthogonal to the $\phi_\mu(s)$, the $g'_{\mu'}$ to the g_μ. The equation $\Gamma_n = h\gamma^n + \Gamma'_n$ shows that $\Gamma'_n/\gamma^n \to 0$, hence $\gamma' < \gamma$.

It is obvious how to continue the process. It yields a string of orthogonal functions $\phi_n(s)$ and orthogonal vectors g_n $(n = 1, 2, \cdots)$ subdivided into sections. The section p, $n_{p-1} < n \leq n_p = n_{p-1} + h_p$, is characterized by a positive eigenvalue γ_p of multiplicity h_p; the vectors g_n of this section form an invariant set and are connected with their partners $\phi_n(s)$ by the relations

$$\sqrt{\gamma_p}\cdot g_n = \int_s \phi_n(s)\cdot sf, \qquad \sqrt{\gamma_p}\cdot\phi_n(s) = (sf, g_n) \qquad [n_{p-1} < n \leq n_p].$$
$$(p = 0, 1, 2, \cdots; \ h_0 = h,\ n_{-1} = 0.)$$

Form the Fourier coefficients $\alpha_n = (g_n, f)$ of f and the remainder

$$f^{(p)} = f - (e + e_1 + \cdots + e_{p-1}) = f - \sum_{k=1}^{n_{p-1}} \alpha_k g_k.$$

Should it ever happen that one of the successive remainders $f, f', \cdots$ vanishes, then the process comes to a stop. But since our final result is trivial in that case, we suppose in our argument that this never comes to pass. γ_p tends to zero, because

$$\sum_{k=1}^{n_p} |\alpha_k|^2 = h\gamma + h_1\gamma_1 + \cdots + h_p\gamma_p \leq \|f\|^2 \leq 1$$

implies

$$(h + h_1 + \cdots + h_p)\cdot\gamma_p = n_p\gamma_p \leq 1.$$

The kernel

$$H^{(p)}(s, t) = (sf^{(p)}, tf^{(p)}) = (sf, tf^{(p)})$$

depends on $t^{-1}s$ only,

$$H^{(p)}(s, t) = H^{(p)}(t^{-1}s), \qquad H^{(p)}(s) = (sf, f^{(p)}).$$

Application of the relation (2.11) to $f^{(p)}$ instead of f gives the inequality $\Gamma_2^{(p)} \leqq \gamma_p \cdot \Gamma_1^{(p)}$ for

$$\Gamma_1^{(p)} = \operatorname{tr}(H^{(p)}) = H^{(p)}(1) = \| f^{(p)} \|^2 \text{ and } \Gamma_2^{(p)} = \int_s | H^{(p)}(s) |^2,$$

and thus

$$\Gamma_2^{(p)} \leq \Gamma_1^{(p)}/n_p. \tag{2.21}$$

By means of the equi-continuity of $H^{(p)}(s) = (sf, f^{(p)})$ for all p we deduce from this estimate for $\Gamma_2^{(p)}$ an estimate for $\Gamma_1^{(p)} = \| f^{(p)} \|^2 = \delta_p^2$ as follows. Choose a positive number ϵ and ascertain a finite number N_ϵ of elements a_i $(i = 1, \cdots, N_\epsilon)$ such that the f-circles C_i of radius ϵ around these elements a_i cover the whole group. We propose to show that $\delta_p \leq 2\epsilon$ as soon as $n_p \geq N_\epsilon/\epsilon^2$. Indeed the contrary hypothesis $2\epsilon < \delta_p$ will lead to $n_p < N_\epsilon/\epsilon^2$. For an element s in the f-circle of radius ϵ around 1 the following inequalities prevail:

$$\| sf - f \| \leq | sf - f | \leq \epsilon,$$

hence

$$| H^{(p)}(s) - H^{(p)}(1) |^2 = | (sf - f, f^{(p)}) |^2 \leq \epsilon^2 \cdot \| f^{(p)} \|^2 = \epsilon^2 \delta_p^2,$$

and finally, since $H^{(p)}(1) = \delta_p^2$ and by hypothesis $\delta_p > 2\epsilon$,

$$| H^{(p)}(s) | \geq \delta_p(\delta_p - \epsilon) > \epsilon\delta_p.$$

Thus $| H^{(p)}(a_i^{-1}s) |^2 > \epsilon^2\delta_p^2$ for $s \,\varepsilon\, C_i$, and consequently

$$\sum_{i=1}^{N_\epsilon} | H^{(p)}(a_i^{-1}s) |^2 > \epsilon^2\delta_p^2$$

everywhere on σ. By integration over s one finds

$$N_\epsilon \cdot \int_s | H^{(p)}(s) |^2 > \epsilon^2\delta_p^2, \quad \text{or} \quad N_\epsilon \cdot \Gamma_2^{(p)} > \epsilon^2\delta_p^2.$$

Combination with (2.21), $n_p\Gamma_2^{(p)} \leq \delta_p^2$, leads to the promised conclusion $n_p < N_\epsilon/\epsilon^2$. But

$$\delta_p^2 = \| f^{(p)} \|^2 = \| f \|^2 - \sum_{k=1}^{n_p-1} | \alpha_k |^2.$$

Hence the result can be stated as follows:

Given any $\epsilon > 0$, *the inequality*

$$(0 \leq) \; \| f \|^2 - \sum_{k=1}^{n-1} | \alpha_k |^2 \leq 4\epsilon^2 \tag{2.22}$$

will hold as soon as $n \geq N_\epsilon/\epsilon^2$.

This not only proves the Parseval equation

$$| \alpha_1 |^2 + | \alpha_2 |^2 + \cdots = \| f \|^2, \tag{2.23}$$

or the fact that the Fourier series $\alpha_1 g_1 + \alpha_2 g_2 + \cdots$ converges weakly to f, but also provides us with an explicit estimate for the remainder under the assumption that the invariant sets are arranged according to descending weights, $\gamma > \gamma_1 > \cdots$.

From f one easily passes to an arbitrary finite linear combination $\sum_i \lambda_i \cdot s_i f$ of transforms of f and any vector g that is the weak limit of a sequence of such combinations. It results that we can approximate every such vector g by a finite sum $v_n = \sum_{k=1}^{n-1} \beta'_k g_k$ in the weak sense, $\| g - v_n \| \leq \epsilon$, with any preassigned accuracy ϵ. But the best approximation in the weak sense by a sum v_n of a given number $n - 1$ of terms is obtained by choosing β'_k as the Fourier coefficients $\beta_k = (g_k, g)$ of g. Hence the Fourier series $\sum \beta_n g_n$ of any g of the type just described, in particular of any vector g in $\Sigma_f{}^0$, converges weakly towards g. In other words, the Parseval equation extends from f to any vector g in $\Sigma_f{}^0$ without a change in the unitary sequence $g_1, g_2, \cdots$ constructed from f.

3. Evaluation of the result. Of the arbitrariness involved in the choice of a unitary basis g_μ $(\mu = 1, \cdots, h)$ in a given invariant set we can make use in such a way that the set breaks up into a number of mutually orthogonal *primitive* sets. Let us do that with each of the sets obtained by our construction! Suppose, for instance, that the first set consists of $h = 8$ vectors and breaks up into two invariant sets of 5 and 3 members respectively. Since the equation (2.15) holds for each vector g_μ of the total set, the relation (2.20) will persist for the two partial sets,

$$\sum_{\mu=1}^{5} | \alpha_\mu |^2 = 5\gamma, \qquad \sum_{\mu=6}^{8} | \alpha_\mu |^2 = 3\gamma.$$

The primitive sets are therefore still arranged according to falling weights, $\gamma \geqq \gamma_1 \geqq \cdots$, however the equality sign can not now be excluded. In the case just mentioned γ_1 would equal γ and 8γ appear as $5\gamma + 3\gamma_1$.

A stage has now been reached where passage from the constructive to the existential standpoint becomes feasible. Let

$$\Omega: \quad s \to \Omega(s) = \| \omega_{\mu\nu}(s) \| , \qquad (\mu, \nu = 1, \cdots, h)$$

be a given irreducible unitary representation of the group σ of degree h. Any representation Ω' that is equivalent to Ω, $\Omega' \sim \Omega$, is also unitary-equivalent to it. [Indeed if the non-singular square matrix A satisfies the condition $\Omega(s)A = A\Omega'(s)$ one sees that AA^* commutes with $\Omega(s)$, and hence by Schur's lemma $AA^* = \rho E$, $\rho > 0$. Here A^* is the conjugate of the transpose of A, and E denotes the unit matrix. $A/\sqrt{\rho}$ is unitary.] Thus we may see to it that, whenever a primitive invariant set obtained in the course of our construction belongs to a representation $\sim\Omega$, its orthogonal basis $g_1, \cdots, g_h$ transforms according *to* Ω *itself*. Schur's lemma further teaches the following two things: (1) If the g_μ and the g'_ν transform according to two *inequivalent* irreducible unitary representations, then they are mutually orthogonal,

$$(g_\mu, g'_\nu) = 0 \qquad (\mu = 1, \cdots, h;\ \nu = 1, \cdots, h').$$

(2) If they transform according to the *same* irreducible unitary Ω, then

$$(g_\mu, g'_\nu) = \beta \cdot \delta_{\mu\nu} \qquad (\mu, \nu = 1, \cdots, h) \tag{3.1}$$

with a factor β independent of μ and ν. Any numerical multiple $\lambda \boldsymbol{g} = (\lambda g_1, \cdots, \lambda g_h)$ of an Ω-row $\boldsymbol{g}$ is an Ω-row, and so is the sum of two such rows. Thus the Ω-rows form a linear manifold. It is natural to introduce the factor β in the equation (3.1) as the scalar product $(\boldsymbol{g}, \boldsymbol{g}')$ of the two Ω-rows $\boldsymbol{g}$ and $\boldsymbol{g}'$;

$$(\boldsymbol{g}, \boldsymbol{g}') = h^{-1} \cdot \sum_\mu (g_\mu, g'_\mu).$$

Indeed it has all the formal properties of a scalar product, in particular $(\boldsymbol{g}, \boldsymbol{g}) > 0$ unless $\boldsymbol{g} = 0$. Call the row $\mathfrak{a} = (\bar{\alpha}_1, \cdots, \bar{\alpha}_h)$ of the conjugates $\alpha_\mu = (f, g_\mu)$ of the Fourier coefficients the *f-component* of the Ω-row $\boldsymbol{g}$ and define

$$(\mathfrak{a}', \mathfrak{a}) = \sum_\mu \alpha'_\mu \bar{\alpha}_\mu, \qquad \| \mathfrak{a} \|^2 = (\mathfrak{a}, \mathfrak{a}).$$

$\lambda\mathfrak{a}$ and $\mathfrak{a} + \mathfrak{a}'$ are the f-components of λg and $g + g'$. The equation (2.19) now reads

$$(\boldsymbol{g}', \mathfrak{F}\boldsymbol{g}) = h^{-1} \cdot (\mathfrak{a}', \mathfrak{a}). \tag{3.2}$$

An Ω-row will be said to be *hidden* or flat if its f-component $\mathfrak{a}$ is zero, and *upright* if it is perpendicular to all hidden Ω-rows. Clearly there can not be more than h linearly independent upright Ω-rows. For let $\boldsymbol{g}^{(1)}, \cdots, \boldsymbol{g}^{(m)}$ be any m such rows and form the linear combination $\boldsymbol{g} = \lambda_1 \boldsymbol{g}^{(1)} + \cdots + \lambda_m \boldsymbol{g}^{(m)}$. Whenever its f-component $\lambda_1 \mathfrak{a}^{(1)} + \cdots + \lambda_m \mathfrak{a}^{(m)}$ vanishes, $\boldsymbol{g}$ itself must be zero, because such a $\boldsymbol{g}$ is at the same time upright and flat, therefore $(\boldsymbol{g}, \boldsymbol{g}) = 0$, $\boldsymbol{g} = 0$.

(2.14) shows that the operator $\mathfrak{F}$ changes an Ω-row $\boldsymbol{g}$ into an Ω-row $\mathfrak{F}\boldsymbol{g}$;

according to (3.2) this $\mathfrak{J}g$ is necessarily upright. Any Ω-row $g = \{g_\mu\}$ obtained by our construction satisfies an equation (2.15) with a positive factor γ, $g_\mu = \gamma^{-1} \cdot \mathfrak{J} g_\mu$, and consequently these Ω-rows themselves are upright: *among all possible Ω-rows the construction automatically selects the upright ones* as those that actually occur in the Parseval equation (2.23). Such a principle of selection is needed, since the linear manifold of *all* Ω-rows may not be of finite nor even of denumerable dimensionality.

Let $g^{(1)}, \cdots, g^{(m)}$ $(m \leq h)$ be an orthogonal basis, $(g^{(i)}, g^{(k)}) = \delta_{ik}$, for the upright Ω-rows. Then their f-components $\alpha^{(1)}, \cdots, \alpha^{(m)}$ are also linearly independent. Moreover we have

$$(g_\mu^{(i)}, g_\nu^{(k)}) = \delta_{ik}\delta_{\mu\nu} \qquad (i, k = 1, \cdots, m;\ \mu, \nu = 1, \cdots, h)$$

and

$$\mathfrak{J} g^{(i)} = \sum_k \gamma_{ki} g^{(k)}$$

where the coefficients

$$\gamma_{ki} = (g^{(k)}, \mathfrak{J} g^{(i)}) = h^{-1}(\alpha^{(k)}, \alpha^{(i)})$$

form a Hermitian matrix. $(g, \mathfrak{J}g)$ for the arbitrary upright Ω-row $g = \sum_i \xi_i g^{(i)}$ is the positive-definite Hermitian form

$$G[\xi] = \sum_{i,k} \bar{\xi}_k \gamma_{ki} \xi_i = h^{-1} \| \sum_i \xi_i \alpha^{(i)} \|^2 \tag{3.3}$$

of the m variables ξ_i. One can therefore ascertain m Ω-rows $\check{g}^{(i)}$ such that $g^{(i)} = \sum_k \gamma_{ki} \check{g}^{(k)}$. Then

$$g_\mu^{(i)} = \int_s \check{\phi}_\mu^{(i)}(s) \cdot sf \quad \text{where} \quad \check{\phi}_\mu^{(i)}(s) = (sf, \check{g}_\mu^{(i)}).$$

[In passing it may be observed that if m has the highest possible value h, then the functions $\omega_{\mu\nu}(s)$ will be f-continuous; for then one can express them as linear combinations of the f-continuous functions $(g_\mu^{(i)}, sf)$ by inverting the relations (2.18),

$$(g_\mu^{(i)}, sf) = \sum_\nu \omega_{\mu\nu}(s) \cdot \alpha_\nu^{(i)} \qquad (\mu, \nu, i = 1, \cdots, h).]$$

The orthogonal basis $g^{(i)}$ could be so chosen that $G[\xi]$ is on principal axes,

$$\gamma_{ki} = 0 \text{ for } k \neq i, \qquad \gamma_{ii} = \gamma_i > 0.$$

Then $\mathfrak{J} g_\mu^{(i)} = \gamma_i \cdot g_\mu^{(i)}$. It is this normalized form which results from the construction of 2. The greatest of the m numbers $\gamma_1, \cdots, \gamma_m$ can be defined as the maximum $\gamma(\Omega)$ of the Hermitian form (3.3) for $\sum_i |\xi_i|^2 \leq 1$.

The contributions

$$(3.4) \qquad e(\Omega) = \sum \alpha_\mu{}^{(i)} g_\mu{}^{(i)} \text{ and } \| e(\Omega) \|^2 = \sum | \alpha_\mu{}^{(i)} |^2 = \sum_i \| \mathfrak{a}^{(i)} \|^2$$

of Ω to f and $\| f \|^2$ are clearly independent of the choice of the orthogonal basis $\mathbf{g}^{(i)}$ for the upright Ω-rows. Hence in the final statement we shall ignore this normalization of $G[\xi]$. Nor do the contributions (3.4) change if Ω is replaced by a unitary-equivalent representation, i. e. if each of the m rows $\mathbf{g}^{(1)}, \cdots, \mathbf{g}^{(m)}$ undergoes the same unitary transformation A. The sum $\sum \| e(\Omega) \|^2$ extending over any set of inequivalent irreducible unitary Ω can not exceed $\| f \|^2$ (Bessel's inequality). It is thus further clear that the construction of $\mathbf{2}$ can not help yielding a *complete* basis $\mathbf{g}^{(i)}$ $(i = 1, \cdots, m)$ of the upright Ω-rows for each Ω. Otherwise the sum (2.23) would fall short of $\| f \|^2$ at least by the contributions of the missing $\mathbf{g}^{(i)}$. We summarize our findings in a preliminary Statement and a Theorem.

STATEMENT. *Starting from a given a. p. vector f, our construction accomplishes the following. For every irreducible unitary representation* Ω *of* σ, *degree h, it determines a complete orthogonal basis* $\mathbf{g}^{(1)}, \cdots, \mathbf{g}^{(m)}$ *of those* Ω*-rows* $\mathbf{g} = (g_1, \cdots, g_h)$ *which are perpendicular to all hidden* Ω*-rows, and it picks out a complete set of inequivalent* Ω*'s that actually occur, i. e. for which* $m > 0$. *One has* $m \leq h$, *and the vectors*

$$g_\mu{}^{(i)} \qquad (i = 1, \cdots, m;\ \mu = 1, \cdots, h)$$

lie in $\Sigma_f{}^0$. [They are even of the special form $\int \phi(s) \cdot sf$ where $\phi(s)$ is not only f-continuous but satisfies a strong Lipschitz condition with respect to f, (2.1).] *The* $g_\mu{}^{(i)}$ *for two inequivalent irreducible unitary representations* Ω *are mutually orthongonal. The contribution of* Ω *to f is the sum* $e(\Omega) = \sum \alpha_\mu{}^{(i)} g_\mu{}^{(i)}$ *formed by means of the Fourier coefficients* $\alpha_\mu{}^{(i)} = (g_\mu{}^{(i)}, f)$, *the contribution of* Ω *to* $\| f \|^2$ *equals* $\| e(\Omega) \|^2 = \sum | \alpha_\mu{}^{(i)} |^2$. *The maximum of* $h^{-1} \sum_\mu | \sum_i \alpha_\mu{}^{(i)} \xi_i |^2$ *for* $\sum_i | \xi_i |^2 \leq 1$ *is introduced as the weight* $\gamma(\Omega)$ *with which* Ω *occurs in f.*

THEOREM. *The sum of the contributions* $\sum \| e(\Omega) \|^2$ *extending over the denumerable sequence of inequivalent irreducible unitary* Ω *actually occurring in f equals* $\| f \|^2$. *An explicit estimate of the convergence can be given if the terms are arranged by descending weights* $\gamma(\Omega)$.

It is perhaps not justified to speak of a completeness relation, since a host of "hidden" invariant sets are left in the dark, for the good reason that they do not contribute to f and $\| f \|^2$.

For the "interpretation by scalar functions on Π" the situation is a little simpler. Here, according to an observation made by E. Cartan [**3**, cf. also **11**], the number of linearly independent Ω-rows is at most h for any irreducible unitary representation Ω of degree h. We choose an orthogonal basis $\boldsymbol{g}^{(1)}, \cdots, \boldsymbol{g}^{(m)}$ for them in order to determine the contribution $e(\Omega)$ to f, *without rejecting the hidden ones.* But even then f must be used for picking out the denumerable sequence of Ω's that actually contribute to f.

A subgroup σ^0 of σ may be treated directly by observing that f is also almost periodic with respect to σ^0. This is a better procedure than by breaking up the σ-invariant sets obtained by our construction into primitive σ^0-invariant sets. For then one would still face the task of getting rid of all but the upright ones.

In the Main Theorem one can readily pass from a single given a. p. vector f to a finite number, or even a denumerable sequence, of such vectors, $f_1, f_2, \cdots$. An Ω-row $\boldsymbol{g}'$ is f-hidden when $(g'_\mu, f) = 0$ $(\mu = 1, \cdots, h)$. For any $v = 1, 2, \cdots$ we construct an orthogonal basis $\boldsymbol{g}^{(v,1)}, \cdots, \boldsymbol{g}^{(v,m_v)}$ for those Ω-rows $\boldsymbol{g}$ which are perpendicular to all f_v-hidden Ω-rows $\boldsymbol{g}'$, $(\boldsymbol{g}', \boldsymbol{g}) = 0$, and moreover satisfy the relations

$$(f_1, g_\mu) = 0, \cdots, (f_{v-1}, g_\mu) = 0 \qquad (\mu = 1, \cdots, h).$$

Of course $m_v \leq h$. If $m_1, m_2, m_3, \cdots$ all vanish then Ω "does not contribute." Otherwise the whole (finite or infinite) sequence of Ω-rows

$$\boldsymbol{g}^{(1,1)}, \cdots, \boldsymbol{g}^{(1,m_1)}; \ \boldsymbol{g}^{(2,1)}, \cdots, \boldsymbol{g}^{(2,m_2)}; \ \cdots \quad \cdots$$

is orthogonal. For any vector f we may determine the Fourier coefficients $\alpha_\mu{}^{(v,i)} = (g_\mu{}^{(v,i)}, f)$ and form

$$\| e_v(f; \Omega) \|^2 = \sum | \alpha_\mu{}^{(v,i)} |^2 \qquad (\mu = 1, \cdots, h; \ i = 1, \cdots, m_v).$$

The contribution of Ω to $\| f \|^2$ is defined by

$$(3.5) \qquad \| e(f; \Omega) \|^2 = \| e_1(f; \Omega) \|^2 + \| e_2(f; \Omega) \|^2 + \cdots,$$

and

$$(3.6) \qquad \sum_{\Omega} \| e(f; \Omega) \|^2 \leq \| f \|^2,$$

the sum extending over a complete set of inequivalent contributing irreducible unitary Ω. For $f = f_v$ the sum (3.5) is finite, $\| e_u(f_v; \Omega) \|^2$ equaling zero for $u = v + 1, v + 2, \cdots$, and *the Bessel inequality* (3.6) *changes into the corresponding equation.* As a result, an orthogonal sequence $g_1, g_2, \cdots$ of vectors, subdivided into sections that form primitive invariant sets, has been found such that the Fourier series $\alpha_1 g_1 + \alpha_2 g_2 + \cdots$ with the coefficients

$\alpha_n = (g_n, f)$ converges weakly to f *for each of the given a. p. vectors* $f = f_1, f_2, \cdots$.

4. Strong approximation. Interpretation by vector functions on Π. A further axiom connecting "length" $|g|$ with "modulus" $\|g\|$ is needed for the transition from weak to strong convergence. Let us study the integral $g_\xi = \int \xi(s) \cdot sg$ in which $\xi(s)$ is any f-continuous numerical function and the vector g of such nature that sg is an f-continuous vector function of s. In the interpretation by scalar functions on Π this relation reads

$$g_\xi(P) = \int_s \xi(s) \cdot g(s^{-1}P),$$

and thus

$$|g_\xi(P)|^2 \leq \int_s |\xi(s)|^2 \cdot \int_s |g(s^{-1}P)|^2.$$

The second factor on the right is independent of P and has been denoted by $\int_P |g(P)|^2 = \|g\|^2$. If $\|\xi\|^2$ stands for $\int_s |\xi(s)|^2$, we therefore arrive at the inequality

$$|g_\xi| = \underset{P}{\text{l. u. b.}} \ |g_\xi(P)| \leq \|\xi\| \cdot \|g\|.$$

The special case suggests the following general axiom:

Axiom IV. Under the conditions specified above the integral $g_\xi = \int \xi(s) \cdot sg$ satisfies the inequality

$$|g_\xi| \leq \|\xi\| \cdot \|g\|. \tag{4.1}$$

I do not deny its arbitrary character; one would wish to reduce it to some simpler assumptions. But two things can be said for it: (1) It holds in any metric vector space Σ provided one identifies $|g|$ with $\|g\|$. Indeed for $g(s) = sg$ the inequality (C$_2$) gives

$$\left\| \int_s \xi(s) \cdot sg \right\|^2 \leq \|\xi\|^2 \cdot \int_s \|sg\|^2 = \|\xi\|^2 \cdot \|g\|^2.$$

(2) It holds for the interpretation by scalar functions on Π.

The following theorem is an immediate consequence of the new axiom.

$\xi(s)$ being an f-continuous function, the Fourier series of the vector $y = \int_s \xi(s) \cdot sf$,

$$\sum_n \eta_n g_n, \qquad \eta_n = (g_n, y),$$

converges strongly towards y. In the sum $\sum_n$ *members of the same set may not be separated.*

Let us indicate the contribution of the first p invariant sets, $e + e_1 + \cdots + e_{p-1}$, by $\{e + \cdots\}_p$ and carry out our calculations for one of the sets, to which the old notations

$$h, \quad g_\mu, \quad \alpha_\mu, \quad e = \sum_\mu \alpha_\mu g_\mu, \quad \omega_{\mu\nu}(s) \qquad (\mu, \nu = 1, \cdots, h)$$

may refer. We have

$$se = \sum_\mu \alpha_\mu \cdot sg_\mu = \sum_{\mu,\nu} \alpha_\mu \; \omega_{\mu\nu}(s) g_\nu = \sum_\nu (g_\nu, sf) \cdot g_\nu, \tag{4.2}$$

therefore

$$\int_s \xi(s) \cdot se = \sum_\nu (g_\nu, y) g_\nu = \sum_\nu \eta_\nu g_\nu$$

and

$$y = \{\sum_\nu \eta_\nu g_\nu + \cdots\}_p + \int_s \xi(s) \cdot sf^{(p)} \tag{4.3}$$

where $f^{(p)}$ denotes the remainder $f - \{e + \cdots\}_p$. The formula (4.2) for se shows that

$$| se - te | \leq \sum_\nu | (g_\nu, sf - tf) | \cdot | g_\nu | \leq \| sf - tf \| \cdot \sum_\nu | g_\nu |$$

and thus

$$| sf^{(p)} - tf^{(p)} | \leq A \cdot | sf - tf |$$

with $A = 1 + \{\sum_\nu | g_\nu | + \cdots\}_p$. Thus $sf^{(p)}$ is an f-continuous vector function of s, and our axiom yields for the remainder in (4.3) the estimate

$$| \int_s \xi(s) \cdot sf^{(p)} | \leq \| \xi \| \cdot \| f^{(p)} \| .$$

But one knows that $\| f^{(p)} \| = \delta_p$ tends to zero with $p \to \infty$.

THEOREM OF STRONG APPROXIMATION. *f can be strongly approximated with arbitrary accuracy by finite sums of the form* $\sum_{k=1}^{n} \alpha'_k g_k$.

The numerical function $| sf - f | = \rho(s)$ is f-continuous since $| \rho(s) - \rho(t) | \leq | sf - tf |$. Choose any $\epsilon > 0$ and let $l(\rho)$ be a non-negative uniformly continuous function of the variable ρ (≥ 0) which vanishes for $\rho \geq \epsilon$ and for which the integral of the f-continuous function $l(\rho(s)) = \lambda(s)$ is 1. Form the difference

$$\int_s \lambda(s) \cdot sf - f = \int_s \lambda(s) (sf - f).$$

Considering that $l(\rho)\cdot\rho \leq \epsilon\cdot l(\rho)$ for $\rho \geq 0$, namely both for $0 \leq \rho < \epsilon$ and for $\rho \geq \epsilon$, one gets

$$|\int_s \lambda(s)\cdot sf - f| \leq \int_s \lambda(s)\cdot|sf - f| \leq \epsilon \int_s \lambda(s) = \epsilon.$$

But the Fourier series of $y = \int_s \lambda(s)\cdot sf$ converges strongly to y, and hence we obtain coefficients $\eta_k = (g_k, y)$ and an $n = n_{p-1}$ such that

$$|y - \sum_{k=1}^{n}{}' \eta_k g_k| \leq \epsilon.$$

The result is a strong approximation of f with the accuracy 2ϵ.

It follows readily that not only f, but every vector in $\Sigma_f{}^0$ can be approximated in the same manner.

It should be noted that the Axiom IV in general does not carry over from the group σ to a subgroup σ^0.

In the interpretation by functions $f(P)$ on Π it seems unnatural to limit oneself to the case where the value of the function is a number; it may itself be a vector with several components. So it is in physics, where the electromagnetic field strength has 6 and the electronic ψ has 2 or 4 components. We may even admit the value $f(P)$ to be a vector in a metric vector space of infinitely many dimensions. Let us therefore now assume that we are given a metric vector space Σ (Axioms I and II) closed in the Hilbert sense, so that any weakly convergent sequence $f_1, f_2, \cdots, \| f_n - f_m \| \to 0$, converges weakly toward a vector f, $\| f_n - f \| \to 0$. (Should the metric space not be closed, we make it so by Cantor's construction.) Suppose, moreover, we are given a point field Π and a realization of the group σ by a transitive group of transformations $P \to sP$ of this point field (Axiom I_0). We study functions f which associate with every point P of Π a vector $f(P)$ in Σ. The transform sf is then to be defined as the vector function associating with P the vector $sf(s^{-1}P)$. But it seems convenient to define a transform $(s,t)f$ for any *pair* of elements (s,t) of σ by ascribing to $(s,t)f$ the value $sf(t^{-1}P)$ at the point P. The pairs (s,t) form the group $\sigma \times \sigma$ in which σ itself is contained as the subgroup of the diagonal pairs (s,s); $sf = (s,s)f$. The restriction of almost-periodicity to be imposed upon f is twofold:

1. Let us say that the element t of σ lies in the f-disk $D_f(b;\epsilon)$ of radius ϵ around b if $\| f(t^{-1}P) - f(b^{-1}P) \| \leq \epsilon$ for all points P. We assume that for every $\epsilon > 0$ the group σ may be covered by a finite number of such f-disks $D_f(b_k;\epsilon)$ $(k = 1, \cdots, n)$ of radius ϵ. Choose a definite point P_0. For an f satisfying this condition the numerical function $\| f(t^{-1}P_0) \|$ of t

and hence the function $\| f(P) \|$ of P is bounded. We define the length $|f|$ of f by

$$|f| = \text{l. u. b.} \| f(P) \| .$$

This length is invariant with respect to all transformations (s, t) of $\sigma \times \sigma$, $|(s,t)f| = |f|$, in particular to all transformations (s, s) of σ. For any two vector functions $f(P)$, $g(P)$ satisfying our condition the scalar product $\psi(P) = (g(P), f(P))$ is an almost periodic numerical function on Π, and we may therefore form the constant $\int\limits_t \psi(t^{-1}P) = \int\limits_P \psi(P)$ and introduce it as the scalar product (g, f). Again invariance prevails: $((s,t)g, (s,t)f) = (g, f)$.

2. Let us say that the element s lies in the f-circle $C_f(a;\epsilon)$ of radius ϵ around a if $\| sf(P) - af(P) \| \leq \epsilon$ for all points P, and now assume that σ can be covered not only by a finite number of f-disks $D_f(b_k;\epsilon)$ $(k = 1, \cdots, n)$, but also by a finite number of f-circles $C_f(a_i;\epsilon)$ $(i = 1, \cdots, m)$ of arbitrarily small radius ϵ. If s lies in $C_f(a_i;\epsilon)$ and t in $D_f(b_k;\epsilon)$ we have

$$\begin{aligned} &\| sf(t^{-1}P) - a_i f(b_k^{-1}P) \| \\ \leq\ &\| sf(t^{-1}P) - a_i f(t^{-1}P) \| + \| a_i f(t^{-1}P) - a_i f(b_k^{-1}P) \| \\ \leq\ &\epsilon + \| f(t^{-1}P) - f(b_k^{-1}P) \| \leq 2\epsilon, \end{aligned}$$

or $|(s,t)f - (a_i, b_k)f| \leq 2\epsilon$. Hence f is almost periodic with respect to the group $\sigma \times \sigma$ of the pairs (s, t) and a fortiori with respect to the subgroup σ of the diagonal pairs (s, s).

All the axioms I — III are satisfied for the space $\sum$ of the almost periodic vector functions f, both with respect to $\sigma \times \sigma$ and to σ. Also the Axiom IV introduced in this section holds. Indeed consider

$$g_\xi(P) = \int\limits_s \int\limits_t \xi(s,t) \cdot sg(t^{-1}P).$$

One finds by applying (C_2) to the (s,t)-means,

$$\| g_\xi(P) \|^2 \leq \int\limits_s \int\limits_t | \xi(s,t) |^2 \cdot \int\limits_s \int\limits_t \| sg(t^{-1}P) \|^2.$$

The second factor on the right equals

$$\int\limits_t \| g(t^{-1}P) \|^2 = \int\limits_P \| g(P) \|^2 = \| g \|^2.$$

Hence $|g_\xi| \leq \| \xi \| \cdot \| g \|$. The same argument applies to a simple mean value of the type

$$g'_\eta(P) = \int\limits_s \eta(s) \cdot sg(s^{-1}P).$$

5. Appendix. On Maak's approach to the theory of almost periodic functions.

1. The marriage problem. Given n distinct objects $\{a\} = a_1, \cdots, a_n$ (boys) and n distinct objects $\{b\} = b_1, \cdots, b_n$ (girls); moreover a scheme of linkage Q_n according to which an a_i and a b_k are either linked (friends) or not linked. I call the number of elements in a set its rank. A set B of girls is said to be associate to a given set A of boys if no boy in A has friends outside the set B. (Then the complement $\{a\} - A$ is in the same sense an associate of $\{b\} - B$.) Question: What is the necessary and sufficient condition that the boys can be paired with the girls in such a fashion that in each of the n pairs the partners are friends? The following basic condition is obviously necessary: A set A of boys has never an associate set B of girls of lesser rank than A. The fundamental combinatorial lemma asserts that this condition is also sufficient [**9**, **4**, **6**].

Proof. I let the girls $b_1, \cdots, b_n$ choose their partners one after the other, and therefore ask b_1 first to make her choice. Suppose she chooses a_3. The choice should be fair to herself, i. e. a_3 and b_1 should be friends. But it should also be fair to the other girls by not making it impossible for them to find partners among their friends, i. e. the linkage scheme Q_{n-1} arising from Q_n by removing a_3 and b_1 should still satisfy the basic condition. Call a set A of boys of rank r (≥ 1) *distinguished* if it has an associate set $B = (b_1, b_{k_2}, \cdots, b_{k_r})$ of girls that is of the same rank r and contains b_1. Then the second postulate requires that there should be no distinguished set A which does not contain a_3, or a_3 must be in the intersection of all distinguished sets.

We make the basic observation that the intersection of two distinguished sets is again distinguished and is not empty. Indeed let A_i be a distinguished set of rank r_i and B_i an associate of the same rank $(i = 1, 2)$. Let $A = A_1 \cap A_2$ be of rank r and $B = B_1 \cap B_2$ of rank s. The rank s is at least 1, since B contains b_1. B is an associate of A, hence $r \leq s$. Moreover $B_1 \cup B_2$ is an associate of $A_1 \cup A_2$; thus $r_1 + r_2 - r \leq r_1 + r_2 - s$, or $r \geq s$. The resulting equation $r = s$ together with $s \geq 1$ proves the point.

Let $A_0 = (a_{i_1}, \cdots, a_{i_m})$ be a distinguished set *of lowest rank* $m \geq 1$ and $(b_1, b_{k_2}, \cdots, b_{k_m})$ its associate of the same rank. A_0 is contained in every distinguished set A (and therefore unique) since $A \cap A_0$ can not be of smaller rank than A_0. At least one of the boys $a_{i_1}, \cdots, a_{i_m}$ in the set A_0 is a friend of b_1; for otherwise A_0 would have an associate set $(b_{k_2}, \cdots, b_{k_m})$ of rank $m - 1$.

Let therefore b_1 pick one of her friends $a_{1'}$ in A_0, and rearrange the sequence $a_1, a_2, \cdots, a_n$ so that her mate $a_{1'}$ assumes the first place. Then the linkage scheme Q_{n-1} of $a_2, \cdots, a_n$ with $b_2, \cdots, b_n$ satisfies the basic condition. By induction the girls thus solve their marriage problem.

2. The objects to which the combinatorial lemma will be applied are pieces σ_i of the group σ, and linkage of a piece σ' and a piece τ' will mean that they have points in common. [Contrary to Maak, we do not forbid the roof-tiles σ_i to overlap.] Let f be an a. p. vector and $\phi(s)$ an f-continuous numerical function, so that for every $\delta > 0$ there is an $\epsilon > 0$ such that $|sf - tf| \leq \epsilon$ implies $|\phi(s) - \phi(t)| \leq \delta$. Then it implies also $|\phi(as) - \phi(at)| \leq \delta$ for every element a. Therefore the function $\phi(s)$ is almost periodic in the sense that for every δ one can cover σ by a finite number of tiles σ_i such that $|\phi(as) - \phi(at)| \leq \delta$ for any s, t in the same piece σ_i and any a. By a simple argument, which I shall not repeat here, Maak infers from this one-sided the two-sided almost-periodicity.

Let us say that s lies in the circlet of diameter $\rho > 0$ around s_0 if

$$|\phi(asb) - \phi(as_0b)| \leq \tfrac{1}{2}\rho \text{ for all } a \text{ and } b.$$

Our statement means that a finite number of elements c_K $(K = 1, \cdots, N)$ may be ascertained such that the circlets σ_K of diameter ρ around the c_K cover the entire group. *Choose the smallest number N of elements c_K satisfying this condition.* We speak of them as a ρ-lattice and of

$$L_\rho = N^{-1} \cdot \sum_k \phi(c_K)$$

as the lattice average of ϕ. (We differ from Maak by using no other domains but circlets. It is a task of fundamentally simpler nature to determine the minimum number of points which have a given property than the minimum number of domains.) Submit the lattice circlets σ_K to an arbitrary left-translation a, $s \rightarrow as$, $\sigma_K \rightarrow \tau_K = a\sigma_K$. The τ_K are circlets of the same diameter ρ. For any r lattice circlets $\sigma_{K_1}, \cdots, \sigma_{K_r}$ it will never happen that those τ_K that are linked with σ_{K_1} or $\sigma_{K_2} \cdots$ or σ_{K_r} are less numerous. For then we could replace the circlets $\sigma_{K_1}, \cdots, \sigma_{K_r}$ by these τ_K that are linked with them, and thereby lower the total number N. Our combinatorial lemma shows that the σ_K and the τ_K may be paired $(\tau_K, \sigma_{K'})$ in such a way that τ_K and $\sigma_{K'}$ overlap.

The centers $d_K = ac_K$ and $c_{K'}$ of two circlets of diameter ρ which overlap satisfy the inequality

$$|\phi(d_K b) - \phi(c_{K'} b)| \leq \rho$$

for all b. Consequently

$$(5.1) \qquad | N^{-1} \sum_K \phi(ac_K b) - N^{-1} \sum_K \phi(c_K b) | \leq \rho$$

for arbitrary a and b, in particular

$$(5.2) \qquad | N^{-1} \sum_K \phi(ac_K) - N^{-1} \sum_K \phi(c_K) | \leq \rho.$$

By the same token the right-translation $s \rightarrow sb$ yields the inequality

$$(5.3) \qquad | N^{-1} \sum_K \phi(c_K b) - N^{-1} \sum_K \phi(c_K) | \leq \rho.$$

Let a_i $(i = 1, \cdots, n)$ be any number of elements and $\alpha_i \geq 0$ corresponding weights of sum 1. The average $A = \sum_i \alpha_i \phi(a_i)$ was said to oscillate by less than ϵ if

$$| \sum_i \alpha_i \phi(aa_i) - A | \leq \epsilon$$

identically in a. Let A, B be two such averages of ϕ oscillating by less than ϵ and δ respectively. We then have

$$(5.4) \qquad | \sum_i \alpha_i \phi(c_K a_i) - A | \leq \epsilon$$

for each c_K of our ρ-lattice. On the other hand (5.3) gives for $b = a_i$:

$$(5.5) \qquad | N^{-1} \sum_K \phi(c_K a_i) - L_\rho | \leq \rho.$$

(5.4) and (5.5) lead to

$$| N^{-1} \sum_{K,i} \alpha_i \phi(c_K a_i) - A | \leq \epsilon, \qquad | N^{-1} \sum_{K,i} \alpha_i \phi(c_K a_i) - L_\rho | \leq \rho,$$
$$| A - L_\rho | \leq \epsilon + \rho.$$

For the same reason

$$| B - L_\rho | \leq \delta + \rho,$$

consequently $| B - A | \leq \epsilon + \delta + 2\rho$. This must be true for every $\rho > 0$, which is impossible unless

$$| B - A | \leq \epsilon + \delta,$$

cf. formula (1.4).

It is this fact on which the existence of the mean value $J = \int \phi(s)$ depends; any average A oscillating by less than ϵ differs from it by not more than ϵ. According to (5.2) and (5.1) the lattice averages

$$L_\rho\langle\phi\rangle = N^{-1} \sum_K \phi(c_K) \text{ and } L_\rho\langle\phi_b\rangle = N^{-1} \sum_K \phi(c_K b)$$

of $\phi(s)$ and $\phi_b(s) = \phi(sb)$ oscillate by less than ρ. Among themselves they differ by not more than ρ, (5.3). This proves that the absolute difference of $\int \phi(sb)$ and $\int \phi(s)$ can not exceed 3ρ, and as ρ is arbitrary, these two mean values must coincide.

INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

[1] S. Bochner and J. von Neumann, "Almost periodic functions in groups," *Transactions of the American Mathematical Society*, vol. 37 (1935), pp. 21-50.

[2] H. Bohr, "Zur Theorie der fastperiodischen Funktionen I," *Acta Mathematica*, vol. 45 (1925), pp. 29-127.

[3] E. Cartan, "Sur la détermination d'un systeme orthogonal complet dans un espace de Riemann symétrique clos," *Rendiconti del Circolo Matematico di Palermo*, vol. 53 (1929), pp. 217-252.

[4] P. Hall, "On representatives of subsets," *The Journal of the London Mathematical Society*, vol. 10 (1935), pp. 26-29.

[5] Anna Hurevitsch, "Unitary representations in Hilbert space of a compact topological group," *Recueil Mathématique (Matematicheskii Sbornik)*, vol. 13 (1943), pp. 79-86.

[6] W. Maak, "Eine neue Definition der fastperiodischen Funktionen" and "Abstrakte fastperiodische Funktionen," *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, vol. 11 (1935), pp. 240-244, and vol. 11 (1936), pp. 367-380.

[7] J. von Neumann, "Almost periodic functions in a group I," *Transactions of the American Mathematical Society*, vol. 36 (1934), pp. 445-492.

[8] F. Peter and H. Weyl, "Die Vollständigkeit der primitiven Darstellungen einer geschlossenen kontinuierlichen Gruppe," *Mathematische Annalen*, vol. 97 (1927), pp. 737-755.

[9] R. Rado, "Bemerkungen zur Kombinatorik im Anschluss an Untersuchungen von Herrn D. König," *Sitzungsberichte der Preussischen Akademie der Wissenschaft*, vol. 32 (1933), p. 68.

[10] H. Weyl, "Integralgleichungen und fastperiodische Funktionen," *Mathematische Annalen*, vol. 97 (1927), pp. 338-356.

[11] H. Weyl, "Harmonics on homogeneous manifolds," *Annals of Mathematics*, vol. 35 (1934), pp. 486-499.

A CRITERION FOR THE NON-DEGENERACY OF THE WAVE EQUATION.*

By PHILIP HARTMAN and AUREL WINTNER.

The results of this paper are similar to those of the first part of [5] but go in another direction. Correspondingly, the knowledge of [5] will not be presupposed.

For large positive t, let $f = f(t)$ be a real-valued, continuous function. By solutions $x = x(t)$ of the linear differential equation

$$x'' + f(t)x = 0 \tag{1}$$

will be meant only real-valued solutions. Such a solution will be called of class (L^2) if

$$\int^{\infty} x^2 dt < \infty. \tag{2}$$

In terms of the (continuous, non-negative) function

$$f^+(t) = \max(f(t), 0), \tag{3}$$

the following theorem will be proved:

(*) *If, as* $t \to \infty$,

$$f(t) = O_R(t^2), \text{ i. e., } f^+(t) = O(t^2), \tag{4}$$

or, more generally, if

$$\int^{t} f^+(s)\,ds = O(t^3), \tag{5}$$

then (1) *cannot have two linearly independent solutions of class* (L^2).

According to Weyl, (1) cannot have two linearly independent solutions of class (L^2) if

$$f(t) = O_R(1), \text{ i. e., } f^+(t) = O(1) \tag{6}$$

(cf. [9], p. 238). No improvement of (6) is known to us from the literature. In view of the particular case (4) of (5), the improvement of (6) to be obtained is of the order t^2.

---

* Received June 1, 1948.

The result is final, in the sense that the t^2 in (4) cannot be improved to $t^{2+\epsilon}$ (for any $\epsilon > 0$). In order to see this, it is sufficient to apply to the general solution of

$$x'' + t^{2+\epsilon}x = 0, \tag{7}$$

an asymptotic formula (cf. [10]), which implies that every solution of (7) is $O(t^{-\frac{1}{2}-\frac{1}{4}\epsilon})$ and, therefore, of class (L^2) (actually, (7) is merely the normal form of Bessel's equation).

The proof of (*) will depend on an adaptation of the method of "arcus variation" (cf. [6], [8], [4]). It consists of an estimate of $N(t)$, where, if $f(t)$ is real-valued and continuous on the half-line $0 \leqq t < \infty$ and $x = x(t)$ denotes any non-trivial solution ($\not\equiv 0$) of (1), the function $N(t)$ is defined as the number of zeros of $x(s)$ on the interval $0 \leqq s \leqq t$. The choice of the solution $x(t) \not\equiv 0$ with reference to which $N(t)$ is estimated is immaterial. In fact, Sturm's separation theorem implies that the N-functions belonging to different non-trivial solutions of (1) cannot differ by more than ± 1 for any value of t. Hence, the assumption, (8), of the following theorem is a property of the coefficient function f of (1) alone.

(**) *If*

$$\liminf_{t\to\infty} N(t)/t^2 < \infty, \tag{8}$$

(for instance, if

$$N(t) = O(t^2) \tag{8 bis}$$

as $t \to \infty$), *then* (1) *cannot have two linearly independent solutions of class* (L^2).

It was shown in [2] that (1) cannot have two linearly independent solutions of class (L^2) if

$$N(t) = O(1). \tag{9}$$

The improvement of (9) to (8 bis) corresponds to that of (6) to (4). The result is again final, in the sense that the t^2 cannot be improved to $t^{2+\epsilon}$ for any $\epsilon > 0$. This follows again from the asymptotic formula for the solutions of (7).

This does not mean, of course, that the assertion of (**) must fail if (8) is violated. In fact, if

$$f(t) = t^2 \log^2 t, \qquad (t > 1), \tag{10}$$

then, on the one hand, the result of [4] implies that

$$\pi N(t) \sim \int^{t} s \log s \, ds \sim \tfrac{1}{2} t^2 \log t$$

and, on the other hand, the asymptotic formula of [10] shows that no non-trivial solution of (1) is of class (L^2) in the case (10).

It turns out that (*) is a corollary of (**). This situation is due to the following

LEMMA. *If $f(t)$, where $0 \leqq t < \infty$, is a real-valued, continuous function and if $f^{+}(t)$ is defined by* (3), *then*

$$N(t) = O(t \int_{0}^{t} f^{+}(s) \, ds)^{\frac{1}{2}} + O(1) \tag{11}$$

holds for the number of zeros of every non-trivial solution of (1).

Since (11) shows that (5) implies (8) [and even (8 bis)], it will be sufficient to prove the Lemma and (**).

Proof of the Lemma. It is clear from (3) and from Sturm's comparison theorem that, if $N^{+}(t)$ belongs to

$$x'' + f^{+}(t) x = 0$$

in the same way as $N(t)$ belongs to (1), then

$$N(t) \leqq N^{+}(t) + 1.$$

This implies that it is sufficient to prove the Lemma under the assumption that $f = f^{+}$, i. e., that

$$f(t) \geqq 0. \tag{12}$$

In other words, the assertion of the Lemma is that (12) is sufficient for

$$N(t) = O(t \int_{0}^{t} f(s) \, ds)^{\frac{1}{2}} + O(1). \tag{13}$$

The truth of the latter implication is a corollary of the following result of Beurling, quoted by Borg [1], p. 1: If $f(t)$ is real-valued and continuous on an interval $a \leqq t \leqq b$, then no solution $x = x(t) \not\equiv 0$ of (1) can have more than one zero on $a \leqq t \leqq b$ unless

$$\int_a^b |f(t)|\,dt > 4(b-a)^{-1}. \tag{14}$$

In order to deduce from this the estimate (13) in the case (12), let $x(t)$, where $0 \leqq t < \infty$, be that solution of (1) determined by the initial conditions $x(0)=0$ and $x'(0)=1$, and let $0=t_0<t_1<\cdots$ denote the zeros of this solution. It can be assumed that this sequence is infinite, since otherwise $N(t)=O(1)$, and so (13) is trivial. Since the solution $x(t)$ has two zeros on the interval $a \leqq t \leqq b$ if $a=t_{n-1}$ and $b=t_n$, where n is any positive integer, it follows that (14) must now hold for every n. In view of (12), this implies that

$$\int_0^{t_n} f(t)\,dt > 4\sum_{k=1}^{n}(t_k-t_{k-1})^{-1}.$$

On the other hand, the harmonic mean of n positive numbers is majorized by their arithmetic mean,

$$\{n^{-1}\sum_{k=1}^{n}(t_k-t_{k-1})^{-1}\}^{-1} \leqq n^{-1}\sum_{k=1}^{n}(t_k-t_{k-1});$$

finally,

$$\sum_{k=1}^{n}(t_k-t_{k-1}) = t_n,$$

since $t_0=0$. The last three formula lines imply that

$$t_n\int_0^{t_n} f(t)\,dt > 4n^2.$$

In view of the definitions of t_n and $N(t)$, the proof of (13) is now complete.

Since this proves the Lemma, only (**) remains to be proved.

Proof of (**). Let $x=x(t)$ and $x=y(t)$ be two linearly independent solutions of (1), where $f(t)$ is any real-valued, continuous function on the half-line $0 \leqq t < \infty$. Since the Wronskian of $x(t)$ and $y(t)$ is a non-vanishing constant, it can be assumed that

$$xy'-x'y=1. \tag{15}$$

In particular, x and y cannot vanish at the same t, and so the substitution

$$x=r\cos\theta, \qquad y=r\sin\theta \tag{16}$$

determines a unique *positive* $r=r(t)$ and, if $0 \leqq (\theta)_{t=0} < 2\pi$, a unique *continuous* $\theta = \theta(t)$ for $0 \leqq t < \infty$. It follows that $\theta(t)$ has a continuous derivative. In addition,

$$\theta'(t) > 0, \tag{17}$$

since (15) and (16) imply that $r^2\theta' = 1$. If the latter identity is written in the form

$$x^2 + y^2 = 1/\theta',$$

and if it is observed that the assertion of (**) is equivalent to the statement that not *both* $x = x(t)$ and $y = y(t)$ are of class (L^2), it follows that (**) will be proved if it is shown that

$$\int_0^\infty dt/\theta'(t) = \infty \tag{18}$$

is implied by the assumption, (8), of (**).

Since $\theta'(t)$ is continuous and positive, its harmonic mean on any interval is majorized by its arithmetical mean; so that, if $0 \leqq a < a + t$,

$$\{t^{-1} \int_a^{a+t} ds/\theta'(s)\}^{-1} \leqq t^{-1} \int_a^{a+t} \theta'(s)\, ds.$$

Since the integral on the right is identical with $\theta(a + t) - \theta(a)$, it follows that

$$\int_a^{a+t} ds/\theta'(s) \geqq t^2/\{\theta(a + t) - \theta(a)\}.$$

If the numerator and denominator of the function on the right of this inequality are divided by $(t + a)^2$, it follows, by letting $t \to \infty$ while a is fixed, that

$$\int_a^\infty ds/\theta'(s) \geqq \limsup_{t\to\infty} t^2/\theta(t). \tag{19}$$

If (18) is false, then the integral on the left of (19) tends to 0 as $a \to \infty$. On the other hand, the upper limit on the right of (19) is non-negative and independent of a. Hence, if (18) is false, the upper limit on the right of (19) must be 0, i. e., the estimate

$$\liminf_{t\to\infty} \theta(t)/t^2 < \infty \tag{20}$$

cannot hold. Consequently, (18) cannot be false if (20) is true.

It follows that the proof of (**) will be complete if it is shown that (8) implies (20).

To this end, recourse must be had to the definition of $N(t)$. In view of (16) and (17), this definition implies that the integral part of the ratio $\theta(t)/\pi$ differs from the integer $N(t)$ by not more than 1. Hence,

$$\theta(t)=\pi N(t)+O(1),$$

and so (20) is implied by (8).

Appendix.

The criteria (*), (**), as well as all other known criteria applying in this direction (cf. [5]), seem to suggest the truth of the following conjecture: If $f(t)$ and $g(t)$ are continuous functions satisfying the inequality

$$f(t)\leqq g(t) \tag{1}$$

for large positive t, and if

$$x''+g(t)x=0 \tag{2}$$

possesses a solution which is not of class (L^2), then

$$x''+f(t)x=0 \tag{3}$$

also has a solution which is not of class (L^2). The object of this Appendix is the construction of an example disproving this conjecture.

If $f(t)=e^{Ct}$, where C is any positive constant, it follows either from asymptotic formulae (cf. [10]) or from explicit integrations, that every solution of (3) is of class (L^2). Hence, it is sufficient to show that there exist functions $g(t)$ which satisfy the inequality

$$e^{Ct}\leqq g(t) \tag{4}$$

(for some $C>0$ and large t) but are such that not every solution of the corresponding differential equation (2) is of class (L^2).

It will be convenient to first choose such a $g(t)$ to be a suitable step-function (in this connection, cf. [7], p. 50) and then to ascertain that a removal of the jumps of this discontinuous $g(t)$ has no effect on the result. Such a step-function on $g(t)$ can even be chosen monotone, as follows:

$$g(t)=k^2 \text{ if } a_{k-1}<t\leqq a_k, \tag{5}$$

where

$$a_k = 2\pi \sum_{j=1}^{k} j^{-1} \text{ if } k = 1, 2, \cdots \text{ and } a_0 = 0; \tag{6}$$

so that $g(t)$ is defined for $0 < t < \infty$.

Since this step-function is monotone, (4) will be proved for continuous t if it is proved for $t = a_k$. In view of (5), this requires the existence of a $C > 0$ satisfying the inequality

$$e^{Ca_k} \leqq k^2.$$

But the existence of such a C is assured by the fact that

$$a_k \sim 2\pi \log k,$$

by (6).

Next, (5) and (6) show that (2) reduces to

$$x'' + k^2 x = 0 \text{ for } a_{k-1} < t \leqq a_k$$

and admits, therefore, the solution

$$x(t) = \cos k(t - a_{k-1}) \text{ for } a_{k-1} < t \leqq a_k. \tag{7}$$

The function $x(t)$ defined by (7) for $0 < t < \infty$ is continuous and has a continuous first derivative, even at the points $t = a_k$. This follows by observing that

$$a_k - a_{k-1} = 2\pi/k,$$

by (6). Furthermore, by (7) and (6),

$$\int_{a_{k-1}}^{a_k} x^2(s)\,ds = \int_0^{2\pi/k} \cos^2 ks\,ds = \pi/k.$$

Hence, from (6),

$$\int_0^t x^2(s)\,ds \sim \tfrac{1}{2}t \text{ as } t \to \infty, \tag{8}$$

and so the (L^2)-condition $\int^{\infty} x^2(s)\,ds < \infty$ is not satisfied.

All that remains to be ascertained is that the jumps of $g(t)$ can be smoothed out without affecting the result. But this follows, for instance, from (8) and the argument applied in the footnote in [3], pp. 396-397.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] G. Borg, "Ueber die Stabilität gewisser Klassen von linearen Differentialgleichungen," *Arkiv für Matematik, Astronomi och Fysik*, vol. 21 A, no. 1 (1944).

[2] P. Hartman, "On differential equations with non-oscillatory eigenfunctions," *Duke Mathematical Journal*, vol. 15 (1948), pp. 697-709.

[3] ———, "On a theorem of Milloux," *American Journal of Mathematics*, vol. 70 (1948), pp. 395-399.

[4] ——— and A. Wintner, "The asymptotic arcus variation of solutions of linear differential equations of second order," *American Journal of Mathematics*, vol. 70 (1948), pp. 1-10.

[5] ——— and A. Wintner, "Criteria of non-degeneracy for the wave equation," *American Journal of Mathematics*, vol. 70 (1948), pp. 295-308.

[6] T. Levi-Civita, "Sur les équations linéaires à coefficients périodiques et sur le moyen mouvement du noeud lunaire," *Annales Scientifiques de l'École Normale Supérieure*, ser. 2, vol. 28 (1911), pp. 325-376.

[7] H. Miloux, "Sur l'équation différentialle $x'' + A(t)x = 0$," *Prace Matematyczno-Fizyczne*, vol. 41 (1934), pp. 39-54.

[8] W. E. Milne, "On the degree of convergence of expansions in an infinite interval," *Transactions of the American Mathematical Society*, vol. 31 (1929), pp. 906-918.

[9] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugenhörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[10] A. Wintner, "On the normalization of characteristic differentials in continuous spectra," *Physical Review*, vol. 72 (1947), pp. 516-517.

ON THE LOCATION OF SPECTRA OF WAVE EQUATIONS.*

By Philip Hartman and Aurel Wintner.

Let $f(t)$ be a real-valued, continuous function defined for large positive s, and let λ be a real parameter. For every fixed λ, only real-valued solutions $x = x(s) \not\equiv 0$ of

$$x'' + (f(s) + \lambda)x = 0 \tag{1}$$

will be considered. Such a solution will be called of class L^2 if

$$\int^{\infty} x^2(s)\,ds < \infty. \tag{2}$$

The following theorem, particular cases of which are often stated (apparently for intuitive reasons) in the physical literature (cf., e.g., [3], p. 72 and p. 83), will be proved:

(I) *If*

$$\limsup_{s\to\infty} f(s) < \infty \tag{3}$$

and if a solution $x = x(s)$ *of*

$$x'' + (f(s) + \lambda_0)x = 0 \tag{4}$$

satisfies

$$x(s) = O(1), \qquad (s \to \infty), \tag{5}$$

then either (2) *holds or* $\lambda = \lambda_0$ *is in the essential spectrum of* (1).

In order to explain the terminology in this assertion, suppose that $f(s)$ is continuous for $0 \leqq s < \infty$ and that there is assigned to (1) an homogeneous boundary condition at $s = 0$, such as $x(0) = 0$ or, more generally,

$$x(0)\cos\alpha + x'(0)\sin\alpha = 0 \tag{6}$$

(so that $x(0) = 0$ corresponds to $\alpha = 0$). The boundary condition (6) belonging to a fixed α determines for (1) a spectrum $S = S(\alpha)$ and a point spectrum $P = P(\alpha)$ if (1) is of the *Grenzpunkt* type, that is, if, for some λ (but then for every λ), not *all* solutions of (1) satisfy (2); cf. [6], p. 238.

---

* Received June 7, 1948.

It is known ([6], p. 238) that assumption (3) is sufficient in order that (1) be of the *Grenzpunkt* type. It is also known ([6], p. 251) that the set consisting of the cluster points of $S(\alpha)$ is independent of the choice of α. What in (I) is referred to as the essential spectrum is this cluster set which, being independent of α, can be denoted by S'. It is contained in $S(\alpha)$, since the latter set is closed. Since $P(\alpha)$ also is contained in $S(\alpha)$, it follows that (I) can be restated as follows:

(I bis) *On the half-line* $0 \leqq s < \infty$, *let* $f(s)$ *be a real-valued, continuous function satisfying* (3) *and let* $x = x(s) \not\equiv 0$ *be a solution of* (4) *satisfying* (5). *In terms of this* $x(s)$, *define* α (mod π) *by* (6). *Then* $\lambda = \lambda_0$ *is in the spectrum* $S(\alpha)$ *determined by* (1) *and* (6).

It remains undecided whether or not (I bis) remains true if its assumption (3) is generalized to the mere requirement that (1) and (6) determine an eigenvalue problem, that is, that (1) be of the *Grenzpunkt* type. In the physical situations occurring in wave mechanics, not only do the potentials satisfy the unilateral restriction (3) but even

$$f(s) = O(1), \qquad (s \rightarrow \infty). \tag{7}$$

An illustration of the content of (I bis), in the particular case (7), is afforded by the case of a continuous, periodic $f(s)$, where $0 \leqq s < \infty$; cf. [5]. Needless to say, the assertion of (I bis) is known in those cases in which $f(s)$ is so "regularly small" for large s that the asymptotic behavior of the general solution of (1), for large s and for arbitrarily fixed λ, can be obtained from asymptotic formulae of standard types.

It should be noted that the two cases admitted in the alternative statement of (I) are not, in general, mutually exclusive. Examples to this effect can be written down by employing general criteria (cf. [8], p. 269 and [4]).

The first of the cases admitted under the alternative of (I) can be ruled out if it is assumed that (5) is satisfied by *every* solution of (4). This is the content of the following theorem:

(II) *If* (3) *is assumed and if* (4) *has two linearly independent solutions satisfying* (5), *then no solution of* (4) *satisfies* (2) *and* $\lambda = \lambda_0$ *is in the essential spectrum of* (1).

Corresponding to the remark following (I bis), it remains undecided whether or not (II) remains true if its assumption (3) is generalized to the mere requirement that (1) be of the *Grenzpunkt* type.

In the proofs, the following facts will be needed:

If $f(s)$ satisfies (3) *and if λ_0 is arbitrary, then, as $s \to \infty$, a solution of* (4).

(i) *cannot be of class* (L^2), *unless it is $o(1)$;*

(ii) *cannot be $o(1)$ unless the first derivative of the solution is $o(1)$;*

(iii) *cannot be $O(1)$ unless the first derivative of the solution is $O(1)$.*

The assertions (i) and (ii) were proved in [7], p. 8, and [1], pp. 324-325, respectively, and a glance at the proof of (ii) shows that (iii) follows by exactly the same argument as (ii).

It may be mentioned that if the unilateral restriction (3) is strengthened to (7), then (4) shows that the respective assumptions, $x(s) = o(1)$ and $x(s) = O(1)$, of (ii) and (iii), imply that $x''(s) = O(1)$ and so the respective assertions, $x'(s) = o(1)$ and $x'(s) = O(1)$, of (ii) and (iii) are contained in Hadamard's standard Tauberian lemma. But his lemma does not apply if his bilateral assumption (7) is relaxed to (3).

Proof of (I bis). For a fixed α in (6), let $x = x(s) \not\equiv 0$ be a solution of (4) and (6). If this $x(s)$ is of class (L^2), then $\lambda = \lambda_0$ is in the point spectrum, $P(\alpha)$, and therefore in the spectrum, $S(\alpha)$. If $x(s)$ is not of class (L^2), suppose that it satisfies (5). It will be proved that $\lambda = \lambda_0$ must then be in the essential spectrum, S'.

It was shown in [2] that (whether or not (3) is satisfied) every λ-value not contained in S' is such that the equation (1) belonging to this λ-value possesses a solution $x = y(t) \not\equiv 0$ which is of class (L^2). Hence, in order to complete the proof of (I bis), it is sufficient to show that if (3) is assumed of $f(s)$, and (5) of a solution $x = x(s) \not\equiv 0$ of (4), and if $x = y(s)$ is a solution of (4) linearly independent of $x(s)$, then $y(s)$ cannot be of class (L^2).

Suppose the contrary. Then, $y(s)$ being of class (L^2), it follows from (i) and (ii) that

$$y(s) = o(1) \text{ and } y'(s) = o(1). \tag{8}$$

On the other hand, (5) and (iii) show that

$$x'(s) = O(1). \tag{9}$$

Since (5), (8) and (9) imply that the Wronskian of $x(s)$ and $y(s)$ is $o(1)$, and since the Wronskian of two solution of (4) is a constant, it follows that the latter constant is 0. But this contradicts the assumption that $x(s)$ and $y(s)$ are linearly independent solutions of (4).

Proof of (II). It was shown in [9], pp. 23-24, that, if (3) is assumed, and if *all* solutions of (4) and their first derivatives satisfy (5) and (9), then $\lambda = \lambda_0$ is in the essential spectrum of (1). Hence, (II) follows from (iii).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman, "The L^2-solutions of linear differential equations of second order," *Duke Mathematical Journal*, vol. 14 (1947), pp. 323-326.

[2] ——— and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[3] H. A. Kramers, *Grundlagen der Quantentheorie*, Leipzig, 1938.

[4] C. R. Putnam, "On the spectra of certain boundary value problems," *American Journal of Mathematics*, vol. 71 (1949), pp. 109-111.

[5] S. Wallach, "The stability of differential equations with periodic coefficients," *Proceedings of the National Academy of Sciences*, vol. 34 (1948), pp. 203-204.

[6] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[7] A. Wintner, "(L^2)-connections between the potential and kinetic energies of linear systems," *American Journal of Mathematics*, vol. 69 (1947), pp. 5-13.

[8] ———, "Asymptotic integrations of the adiabatic oscillator," *ibid.*, vol. 69 (1947), pp. 251-272.

[9] ———, "On the location of continuous spectra," *ibid.*, vol. 70 (1948), pp. 22-30.

CONTINUOUS DECOMPOSITIONS.*

By G. T. Whyburn.

1. Introduction. The decomposition of a region in the complex plane into sets $f^{-1}(w)$ effected by a single valued function $f(z)$ analytic in this region is lower semi-continuous but not necessarily upper semi-continuous [1, 2, 3, 4, 5] [1] in the stronger or open set sense. In other words, the sum (union) of all elements *intersecting* any open set is open but the sum of all elements *contained in* an open set may be neither empty nor open. Indeed it is readily seen that if we take the region to be the whole plane, no entire transcendental function can generate a decomposition upper semi-continuous in this sense. Of course the decomposition generated by a rational function is u. s. c. in this sense, because the function is extensible to the compact plane (= complex sphere) and, for compact spaces, open set upper semi-continuity is equivalent to the weaker limit sense upper semi-continuity which always holds for the decomposition generated by any continuous mapping.

Since the stronger open set continuity of a decomposition usually limits the decomposition to one having compact elements, whereas the transcendental entire functions *never* yield such a decomposition, it seems desirable to develop a theory for decompositions into non-compact elements. Further, there is need for the introduction of a weaker type of continuity which will hold for the decompositions generated by significant classes of analytic functions and from which important consequences can be drawn. This in brief is the objective of the present paper.

The term *topological space* will be used in the Hausdorff sense, i. e., it is a space satisfying Hausdorff's three fundamental neighborhood or open set axioms [6] plus the weakest separation axiom (for any two distinct points x, y there is an open set containing x but not y). When we have a metric space, distances will be denoted by $\rho(x, y)$ and circular neighborhoods with center x and radius r by $V_r(x)$.

A *mapping* $f(A) = B$ is a single valued continuous transformation of A onto B. Such a mapping is *interior* [7, 8] or *open* provided the image of every open set in A is open in B, and *closed* if the image of every closed set in A is closed in B.

Let S be a topological space and let G be a decomposition or partitioning

---

* Received June 11, 1948.

[1] Numbers in brackets refer to the bibliography at the end of the paper.

of S into disjoint closed sets. For any set X in S let X_g denote the sum of all elements of G intersecting X. We consider the conditions:

(a) *The sum of all elements of G intersecting an open set is open* (i. e., U open implies U_g open)

(b) *The sum of all elements of G intersecting a closed set is closed* (F closed implies F_g closed)

(b′) *The sum of all elements of C intersecting a compact set is closed* (K compact implies K_g closed)

(b″) *Any element of G intersecting the limit inferior of a sequence of elements of G contains the limit superior of this sequence.*

(1. 1) THEOREM. *If S is weakly separable,*[2] *then* (b′) *and* (b″) *are equivalent. In particular,* (b′) *and* (b″) *are equivalent when S is perfectly separable.*

Proof. Suppose (b′). Then let $g_1, g_2, \cdots$ be a sequence with $x \varepsilon g \varepsilon G$ and $x \varepsilon \underline{\lim}\, g_n$. Let $y \varepsilon \overline{\lim}\, g_n$. It follows from our hypothesis on S that there exists a sequence (x_n) with $x_n \varepsilon g_n$ and so that $x_n \rightarrow x$ in the sense that every open set about x contains almost all the points x_n. Then if $X = x + \sum_r^\infty x_n$, where r is chosen so that y non $\varepsilon\, g_n$ for $n \geqq r$, X is compact. Accordingly X_g is closed. Thus $y \varepsilon X_g$; and since y non $\varepsilon\, g_n$, $n \geqq r$, and $X_g = g + \sum_r^\infty g_n$, we must have $y \varepsilon g$. Whence $g \supset \overline{\lim}\, g_n$.

Suppose, on the other hand, that (b″) holds and let K be any compact set in S. If x were a limit point of K_g not belonging to K_g, we could find an infinite sequence of distinct points $x_1, x_2, \cdots$ in K_g converging to x. Then if $g_1, g_2, \cdots$ are the elements of G containing $x_1, x_2, \cdots$ respectively, $\underline{\lim}\, g_n \supset x$. However, the element g of G containing x does not intersect K (since x non $\varepsilon\, K_g$), whereas $\overline{\lim}\, g_n \cdot K \neq 0$ since K is compact and $K \cdot g_n \neq 0$ for all n.

2. Semi-closed mappings. A mapping $f(A) = B$ will be called *semi-closed* provided the image of every compact set in A is closed in B, it being understood that A and B are topological spaces. Of course, the image of every compact set is compact but it may fail to be closed unless further restrictions are put on the image space B or on the mapping. This is brought out sharply in the following:

---

[2] That is, satisfies the first Hausdorff countability axiom that for each $p \varepsilon S$ there exists a monotone decreasing sequence $U_1 \supset U_2 \supset \cdots$ of open sets closing down on p in the sense that for any open set U in S containing p, we have $p \subset U_n \subset U$ for some n.

EXAMPLE. *There exists an interior mapping* $f(A)=B$ *of a compact metric space* A *onto a compact and countable but non-regular topological space* B.

For let A consist of two disjoint sequences of real numbers (x_n) and (y_n) converging to distinct limits a and b with the real number system metric. Let B consist of points a', b', z_1, z_2, $z_3, \cdots$; and in B let the empty set, sets of the types (z_n), $a'+\sum_{n}^{\infty} z_i$, $b'+\sum_{n}^{\infty} z_i$ (for all n), and any set which is the sum of sets of this type be open sets. It is then readily verified that A is compact and metric and B is a topological space. Also B is compact since every infinite set of points in B has both a' and b' as limit points. However, B is not regular, since the closure of every open set containing a' also contains b'. Finally, if we define $f(x_n)=f(y_n)=z_n$ $(n=1,2,\cdots)$, $f(a)=a'$, $f(b)=b'$, f maps A continuously and interiorly onto B. This mapping is not semi-closed, because the compact set $a+\sum_{1}^{\infty} x_n$ maps onto the set $a'+\sum_{1}^{\infty} z_n$ which is compact but not closed as b' is a limit point of it.

(2.1) THEOREM. *Let* $f(A)=B$ *be an interior and semi-closed mapping, where* A *and* B *are topological spaces. If* A *is perfectly separable, so also is* B. *If* A *is regular and locally compact, so also is* B. *If* A *is locally compact, separable and metric, so also is* B.

Proof. To prove the first statement, let $R_1, R_2, \cdots$ be a fundamental sequence of open sets (basis) in A. Then $f(R_1), f, R_2), \cdots$ is a fundamental sequence of open sets in B. For if $x \varepsilon B$ and U is any open set in B containing x, $f^{-1}(U)$ is open and hence contains a set R_m which intersects $f^{-1}(x)$. Accordingly $x \varepsilon f(R_m) \subset U$.

For the second statement, let $x \varepsilon B$ and let U be any open set in B containing x. Since $f^{-1}(U)$ is open, there exists an open set V such that $V \cdot f^{-1}(x) \neq 0$ and $\bar{V}$ is compact and contained in $f^{-1}(U)$. Then $f(V)$ is open, and contains x and $\overline{f(V)} \subset f(\bar{V}) \subset U$, since $f(V) \subset f(\bar{V})$ and $f(\bar{V})$ is closed.

The final statement is an immediate consequence of the two preceding ones.

(2.11) COROLLARY. *Given* $f(A)=B$ *interior, where* A *and* B *are topological spaces, if* A *is perfectly separable, so also is* B.

In any weakly separable topological space satisfying the axiom that two distinct points lie in disjoint open sets, every compact set is closed. Thus in particular in a regular topological space every compact set is closed. Whence

(2.2) THEOREM. *Any continuous mapping $f(A) = B$ of a topological space A onto a weakly separable regular topological space B is semi-closed.*

(2.3) THEOREM. *If A is locally compact, separable and metric and $f(A) = B$ is interior, where B is a topological space, a necessary and sufficient condition that B be locally compact, separable and metric is that f be semi-closed.*

Proof. By the above theorem, separability and regularity of B imply that f is semi-closed. On the other hand, if f is semi-closed, by (2.1) B is locally compact, separable and metric.

3. The decomposition space. Natural mapping. Let G be a decomposition of a topological space S into disjoint closed sets. We set up a decomposition or hyperspace S' as follows. The elements of G become the points of S', and a set in S' is open if and only if the sum (union) of the corresponding elements of G is an open set in S. It follows at once that if G satisfies (a) a set in S' is open if and only if the sum of the corresponding elements of G is the set X_g for some *open* set X in S. For convenience we also define any open set S' containing $x' \varepsilon S'$ to be a neighborhood of x'.

For convenience of reference we state next the well known

(3.1) THEOREM. *The decomposition space S' is a topological space.*

For a proof of this theorem the reader is referred to [2], p. 61.

Definition. If for each $x \varepsilon S$ we define $\phi(x)$ to be the point in S' given by the element of G containing x, we obtain the *natural mapping* of S onto S' generated by G.

It follows directly from the definition that this mapping ϕ is continuous if condition (a) is satisfied, as is well known. Also, in this case, we have

(3.2) THEOREM. *The natural mapping ϕ generated by a decomposition G satisfying condition* (a) *is interior.*

For, if U is any open set in S, $\phi(U)$ is the set of points in S' given by elements of G intersecting U and thus $\phi(U)$ is open in S'. Thus from (2.11) we have

(3.21) COROLLARY. *If S is perfectly separable so also is S'.*

(3.3) THEOREM. *The natural mapping ϕ generated by a decomposition G of S satisfying conditions* (a) *and* (b′) *is also semi-closed.*

For let K be any compact set in S and let $H = \phi(K)$. Then if K_g

denotes the sum of all elements of G intersecting K, $S - K_g$ is open (since K_g is closed). Accordingly $\phi(S - K_g) = S' - H$ is open, so that H is closed.

Thus from (2.1) we have

(3.4) THEOREM. *If the decomposition G of a locally compact separable metric space S satisfies* (a) *and* (b′); *the hyperspace S' is locally compact separable and metric and the natural mapping $\phi(S) = S'$ is continuous, interior and semi-closed.*

4. Decompositions generated by mappings. We next consider the converse situation in which we start with an interior mapping and study the natural decomposition generated in the range space.

(4.1) THEOREM. *Let $f(A) = B$ be an interior mapping, where A and B are topological spaces, let G be the decomposition of A into the sets $[f^{-1}(y)]_{y \varepsilon B}$, let A' be the hyperspace of this decomposition and $\phi(A) = A'$ the associated natural mapping. Then ϕ is interior. Further, the mapping $f\phi^{-1}$ is $1 - 1$ and interior and thus is a homeomorphism of A' onto B.*

To show that ϕ is interior it suffices in view of (3.2) to show that the decomposition G satisfies condition (a). To this end let U be any open set in A. Then since U_g is precisely the set $f^{-1}f(U)$, it follows that U_g is open since $f(U)$ is open by interiority of f and $f^{-1}(U)$ is open by continuity of f.

To see that $f\phi^{-1}$ is interior, let V be any open set in A'. Then $\phi^{-1}(V)$ is open by continuity of ϕ and $f\phi^{-1}(V)$ is then open in B by interiority of f.

(4.11) COROLLARY. *If B is regular and weakly separable, so also is A'; and the mappings f and ϕ are semi-closed.*

Now if we denote the homeomorphism $f\phi^{-1}$ by h, it results at once that for each $x \varepsilon A$

$$h\phi(x) = f\phi^{-1}\phi(x) = f(x).$$

Whence,

(4.2). *The natural mapping ϕ generated by the decomposition given by an interior mapping $f(A) = B$ is topologically equivalent to f.*

Thus the cycle closes when we begin with an interior mapping, take the decomposition of the range space, set up the decomposition space and take the natural mapping of the decomposition, in the sense that we get back a mapping topologically equivalent to the one we started with. This is not true in general with merely continuous mappings on non-compact spaces.

Inasmuch as decompositions generate natural mappings of the original space onto the hyperspace, and mappings generate natural decompositions of the original space, and these two processes are equivalent in a very real sense, it would be possible to dispense with the study of decompositions and just study mappings or conversely. However, both viewpoints have proven most fruitful in the past (See [1, 2, 3, 4, 8]). Each has contributed materially to the development and enrichment of topological and function-theoretic results, due to the fact that some relationships reveal themselves intuitively in terms of decompositions while others show up more naturally as mapping theorems. Thus the two viewpoints have played a complementary role in appealing to mathematical intuition; and it seems reasonable to suppose they may continue to do so.

5. Upper semi-continuity in the open set sense. A decomposition G of a topological space S into disjoint closed sets satisfying condition (b) is u. s. c. in the open set sense. As originally defined by Moore [1], Alexandroff [2] and others, G is u. s. c. provided that if $g \varepsilon G$ and U is any open set about g, there exists an open set V about g such that any element of G intersecting V lies in U. Clearly the two definitions are equivalent to each other and equivalent in turn to the condition that the sum of all elements of G contained in any given open set be open (possibly empty). It is clear that in a compact metric space, condition (b), (b′) and (b″) are equivalent. Also [8] in a locally compact separable metric space, (b) and (b′) are equivalent provided the elements of G are continua.

That condition (b), in the presence of (a), tends very strongly to limit the decomposition to having compact elements and thus is too strong for our purposes is shown by

(5. 1) THEOREM. *If G is a non-degenerate decomposition of a connected separable metric space into disjoint closed sets satisfying conditions* (a) *and* (b), *the elements of G are necessarily compact.*

Proof. This theorem follows from a theorem of A. D. Wallace [9]. For it is readily verified that the conditions in Wallace's theorem follow from conditions (a) and (b).

6. Continuity condition for locally connected spaces. We next consider the following conditions for a decomposition G of a topological space S satisfying (a).

(c) *For any open set U in S which is the sum of the elements of a subcollection of G corresponding to a connected set in S' and any component Q of U, we have $Q_g = U$.*

(c′) *For any region R in the decomposition space S', each component of the inverse of R maps onto R under the natural mapping ϕ.*

(c″) *For any* $y \varepsilon S'$ *and any open set* U *in* S' *containing* y, *there exists a region* R *with* $y \subset R \subset U$ *such that each component of the inverse of* R *maps onto* R *under* ϕ.

Note. A *region* is a connected open set.

It is clear that (c′) is merely a restatement of (c) in terms of the natural mapping. Hence (c) and (c′) are equivalent in any topological space.

(6.1) THEOREM. *If* S *is a locally connected topological space, then for any decomposition* G *of* S *into disjoint closed sets satisfying* (a), *conditions* (c), (c′) *and* (c″) *are equivalent.*

It suffices to show that (c′) and (c″) are equivalent; and since (c′) obviously implies (c″) we have left to show that (c″) implies (c′).

To this end let R be any region in B and Q any component of $\phi^{-1}(R)$. Since $\phi(Q)$ is open and $\phi(Q) \subset R$, if $\phi(Q) \neq R$ there would exist a point $y \varepsilon \overline{\phi(Q)}[R - \phi(Q)]$. Let E be a region satisfying $y \subset E \subset R$ and such that any component of $\phi^{-1}(E)$ maps onto E under ϕ. Since $\overline{\phi(Q)} \supset y$, there exists a $z \varepsilon \phi(Q) \cdot E$. Let $w \varepsilon Q \cdot \phi^{-1}(E)$ and let H be the component of $\phi^{-1}(E)$ containing w. By hypothesis $\phi(H) = E \supset y$. But since $\phi(H) = E \subset R$, we have $H \subset \phi^{-1}(R)$; and since $H \cdot Q \supset w \neq 0$ and Q is a component of $\phi^{-1}(R)$, we must have $Q \supset H$ and therefore $\phi(Q) \supset \phi(H) = E \supset y$, contrary to $y \varepsilon R - \phi(Q)$.

7. Continuity in the open set sense. A decomposition G of a topological space S into disjoint closed sets satisfying (a) and (b) will be called continuous in the open set sense. That is, G is continuous in the open set sense if for any open set U in S the sums U_g of all elements intersecting U and U_0 of all elements contained in U are open sets.

(7.1) THEOREM. *In a locally connected compact separable metric space* S, *any decomposition* G *of* S *satisfying* (a), (b′) *and* (c) *is upper semi-continuous in the open set sense at each compact element of* G.

This means that if $X \varepsilon G$ is compact and U is any open set containing X, there exists an open set V such that any element of G intersecting V lies in U. By local compactness of S we may suppose U so chosen that $\bar{U}$ is compact. We shall now show that the sum U_0 of all elements of G contained wholly in U is an open set so that it may be taken as our V. To this end let Y be any element of G contained in U. Let $F(U)$ denote the boundary $\bar{U} - U$ of U, let ϕ be the natural mapping of G and let R' be the component of $S' - \phi[F(U)]$ containing $\phi(Y)$. Since each component of $\phi^{-1}(R')$ intersects Y, by (c), whereas $\phi^{-1}(R') \cdot F(U) = 0$, we have

$$\phi^{-1}(R') \subset U.$$

Since $\phi^{-1}(R')$ is open and contains Y, it follows that the sum U_0 of all elements of G lying wholly in U is open.

(7.2) THEOREM. *In order that a non-degenerate decomposition G of a connected, locally connected, locally compact separable metric space S into disjoint closed sets be continuous in the open set sense, it is necessary and sufficient that the elements of G be compact and satisfy conditions* (a), (b′) *and* (c).

To prove the necessity, suppose G is continuous in the open set sense, i. e., that it satisfies (a) and (b). Then by (5.1) the elements of G are compact. Condition (b′) follows from (b) by (1.1) since (b) implies (b″) in any separable metric space. To show (c), by (6.1) it suffices to establish (c″). To this end let $y \in S'$ and let V be any open set in S containing $\phi^{-1}(y)$ and such that $\bar{V}$ is compact. By (b) the sum V_0 of all elements of G contained in V is open and hence $\phi(V_0)$ is open by interiority of ϕ. Now if U is any given open set in S' containing y we have $y \in U \cdot \phi(V_0)$; and if R is any region in S' satisfying $y \subset R \subset U \cdot \phi(V_0)$ and Q is any component of $\phi^{-1}(R)$, we shall show that $\phi(Q) = R$. This results from the fact that $Q \subset V_0 \subset V$ and hence Q is conditionally compact. For if $\phi(Q) \neq R$, since $\phi(Q)$ is open, there would be a limit point p of $\phi(Q)$ belonging to $R - \phi(Q)$. Then if $p_1, p_2, \cdots$ is a sequence in $\phi(Q)$ converging to p and $x_n \in Q \cdot \phi^{-1}(p_n)$, there would exist a limit point x of (x_n) in $\bar{Q}$. But by continuity of ϕ, $\phi(x) = p$ so that $x \in \phi^{-1}(p) \subset \phi^{-1}(R)$; and this gives $x \in Q$ since Q is a component of $\phi^{-1}(R)$. Accordingly $\phi(Q) = R$ and (c) is established.

The sufficiency of the conditions is a direct consequence of (7.1).

In connection with (7.2) it may be of interest to remark that in case A is compact (i. e., A is a locally connected continuum) compactness of the point inverses results from the continuity of f and condition (b′) is equivalent to (b). Thus (a) and (b′) are equivalent to (a) and (b), and we obtain (c) as a necessary consequence of (a, b). This latter fact is known (See (7.4), p. 147 of [8]). The property (c) may be used to characterize the quasi-monotone mappings on locally connected continua, as shown by Wallace (See p. 152 of [8] for example).

8. Conclusion. Our results indicate that in spaces which are locally connected generalized continua (= locally compact, separable, metric and connected), conditions (a), (b′) and (c) provide limitations on a decomposition which it would be reasonable to call *continuity* of the decomposition. Surely open set continuity is too strong if we wish to make more effective contact with analytic functions. For it is now clear that the decomposition generated in the complex plane by no transcendental entire function can be continuous in the open set sense, since this would imply compactness of the elements whereas at most one of them could be compact by the Picard Theorem. Further, in case the elements of the decomposition are compact, conditions (a), (b′) and (c) reduce to open set continuity as shown in (7.2).

On the other hand, the decomposition effected in the complex plane by many transcendental entire functions will satisfy these conditions. For example the exponential function e^z, $\sin z$, or indeed any periodic entire function, generates a decomposition continuous in this (a, b′, c) sense. Also it will be shown in another paper that the decomposition in the complex plane effected by any entire function of order $< 1/2$ is continuous in this sense. In the same paper significant consequences will be established concerning mappings which generate (a, b′, c) continuous decompositions. For example, a function such as ze^z or $(z - a_1), (z - a_2) \cdots (z - a_k) e^z$ would not generate a continuous decomposition in this sense because these functions vanish at only a finite number of points but take other values infinitely many times.

UNIVERSITY OF VIRGINIA.

BIBLIOGRAPHY.

1. R. L. Moore, *Foundations of Point Set Theory*, American Mathematical Society Colloquium Publications, vol. 13 (1932). See references therein to earlier papers on semi-continuous collections by Moore and others.
2. P. Alexandroff and H. Hopf, *Topologie I*, Berlin, 1935, Springer.
3. L. Vietoris, "Über stetige Abbildungen einer Kugelfläche," *Proceedings, Akademie van Wetenschappen*, Amsterdam, vol. 29 (1926), pp. 443-453.
4. C. Kuratowski, "Sur les décompositions semi-continues d'espaces métriques compacts," *Fundamenta Mathematica*, vol. 11 (1928), pp. 169-185.
5. R. Vaidyanathaswamy, *Treatise on Set Topology*, 1947, Indian Mathematical Society.
6. F. Hausdorff, *Mengenlehre*, 1927, Berlin and Leipzig.
7. S. Stoilow, *Leçons sur les Principes Topologiques de la Théorie des Fonctions Analytiques*, 1938, Paris.
8. G. T. Whyburn, *Analytical Topology*, American Mathematical Society Colloquium Publications, vol. 28 (1942).
9. A. D. Wallace, "Some characterizations of interior transformations," *American Journal of Mathematics*, vol. 61 (1939), p. 761.

QUASI-MONOTONE SERIES.*

By Tomlinson Fort.

Monotone series, that is, series of the type $a_1 + a_2 + \cdots + a_n + \cdots$ where $a_n \geqq a_{n+1} > 0$, have been extensively studied. Szász [this Journal, vol. 70 (1948), p. 203] generalizes the idea of monotoneity to series with

$$(1) \qquad 0 < a_{n+1} \leqq (1+\alpha/n)a_n, \qquad \alpha \geqq 0.$$

For series of this type he establishes the "Cauchy condensation test" and the "Cauchy integral test" for convergence. In the present note the idea of monotoneity is further generalized not only in the domain of reals but also in the domain of complex numbers.

We consider the series

$$(2) \qquad a_1 + a_2 + \cdots + a_n + \cdots,$$

where $a_n = x_n + y_n i$. We assume $x_n \geqq 0$, $y_n \geqq 0$ and let $t_n = x_n + y_n$. We then assume

$$(3) \qquad t_{n-2} + t_{n-1} \geqq 2t_n(1 - 1/b_n)$$

and

$$(4) \qquad 2(1 + 1/B_{n-2})t_{n-2} \geqq t_{n-1} + t_n,$$

where

$$1 < b_n \leqq b_{n+1} \text{ and } 0 < B_n \leqq B_{n+1}.$$

Under these circumstances[1] we call (2) *quasi-monotone in the mean.*

1. We prove the following lemma:

Lemma I. $t_n + t_{n-1} + t_{n-2} + \cdots + t_m + t_{m-1}$

$$(5) \qquad \geqq t_n[1 + (1 - 1/b_n) + (1 - 1/b_n)(1 - 1/b_{n-1}) + \cdots + (1 - 1/b_n)(1 - 1/b_{n-1}) \cdots (1 - 1/b_{m+1})], \quad n > m.$$

---

* Received June 28, 1948.

[1] Relations (3) and (4) can easily be generalized so as to include a larger number of terms. It is also true that for certain results it is not necessary to assume both (3) and (4). This is true of Theorem I.

Proof of Lemma I. We establish the following relation which holds either for $\delta = 0$ or for $\delta = 1$.

$$\begin{aligned} & t_n + t_{n-} + t_{n-2} + \cdots + t_m + \delta t_{m-1} \\ (6) \qquad & \geqq t_n[1 + (1 - 1/b_n) + (1 + 1/b_n)(1 - 1/b_{n-1}) + \cdots \\ & \quad + (1 - 1/b_n)(1 - 1/b_{n-1}) \cdots (1 - 1/b_{m+1}) \\ & \quad + \delta(1 - 1/b_n)(1 - 1/b_{n-1}) \cdots (1 - 1/b_m)]. \end{aligned}$$

Relation (3) implies

$$t_{n-2} + t_{n-1} \geqq (1 - 1/b_n)t_n + (1 - 1/b_n)(1 - 1/b_{n-1})t_n.$$

If $t_{n-1} \geqq (1 - 1/b_n)t_n$, we write

$$\begin{aligned} (7) \qquad t_n + t_{n-1} + \cdots + t_m + \delta t_{m-1} \geqq t_n[1 + (1 - 1/b_n)] + t_{n-2} \\ + \cdots + t_m + \delta t_{m-1}. \end{aligned}$$

If $t_{n-1} < (1 - 1/b_n)t_n$, then necessarily

$$(8) \qquad t_{n-2} > (1 - 1/b_n)(1 - 1/b_{n-1})t_n$$

and we write

$$\begin{aligned} & t_n + t_{n-1} + \cdots + \delta t_{m-1} \\ (9) \qquad & \geqq t_n[1 + (1 - 1/b_n) + (1 - 1/b_n)(1 - 1/b_{n-1})] \\ & \quad + t_{n-3} + t_{n-4} + \cdots + t_n + \delta t_{m-1}. \end{aligned}$$

Next substitute in (7) for t_{n-2} the expression $(1 - 1/b_{n-1})t_{n-1}$ if $t_{n-2} \geqq (1 - 1/b_{n-1})t_{n-1}$ and then for t_{n-1} as above. If $t_{n-2} < (1 - 1/b_{n-1})t_{n-1}$, we substitute for $t_{n-2} + t_{n-3}$ the expression

$$[(1 - 1/b_{n-1}) + (1 - 1/b_{n-1})(1 - 1/b_{n-2})]t_{n-1}$$

and then $(1 - 1/b_n)t_n$ for $t_{n\ 1}$. We substitute similarly in (9) for t_{n-3} or $t_{n-3} + t_{n-4}$ and then for t_{n-2} from (8). The value 1 or 0 is assigned to the symbol δ according as the term t_{m-1} can or can not be used in the last step of the process.

Let $\delta = 1$ in the left-hand member of (6) and $\delta = 0$ in the right-hand member and we have (5).

We now state without proof the following lemma.

Lemma II. $t_m + t_{m+1} + \cdots + t_{n-1}$

$$\begin{aligned} & \leqq t_m[1 + (1 + 1/B_m) + (1 + 1/B_m)(1 + 1/B_{m+1}) \\ & + \cdots + (1 + 1/B_m)(1 + 1/B_{m+1}) \cdots (1 + 1/B_{n-1})], \quad m < n. \end{aligned}$$

THEOREM I. *Let* (2) *be quasi-monotone in the mean and convergent. Let m be an integer less than n but a function of n such that*

$$m \to \infty \text{ when } n \to \infty \text{ and } [(n-m)/b_m] > \eta > 0.$$

Then $a_n b_m \to 0$.

Interesting special cases of this theorem are: $b_n = n^{\alpha}$, $0 < \alpha \leqq 1$. $m = [n/2]$.[2] We then have $n^{\alpha} a_n \to 0$. Similarly, if $b_n = \log n$ we see that $(\log n) a_n \to 0$.

Proof of Theorem I. Choose M so that when $n > M$

$$| a_{m-1} + a_m + \cdots + a_n | \leqq \epsilon.$$

Then, using Lemma I,

$$\begin{aligned} 2^{\frac{1}{2}}\epsilon &\geqq t_{m-1} + t_m + \cdots + t_n \\ &\geqq t_n[1 + (1-1/b_n) + (1-1/b_n)(1-1/b_{n-1}) \\ &\qquad + \cdots + (1-1/b_n)(1-1/b_{n-1}) \cdots (1-1/b_{m+1})] \\ &\geqq t_n[1 + (1-1/b_m) + (1-1/b_m)^2 + \cdots + (1-1/b_m)^{n-m}] \\ &= t_n b_m[1-(1-1/b_m)^{n-m+1}] = t_n b_m[1-\{(1-1/b_m)^{b_m}\}^{(n-m+1)/b_m}] \\ &= t_n b_m \rho_n, \text{ where } \rho_n > \zeta > 0. \end{aligned}$$

Hence $b_m t_n \to 0$. But $t_n \geqq | a_n |$. Hence $b_m a_n \to 0$.

Next let us be given a sequence of integers g_k such that

$$(10) \qquad g_{k+2} - g_{k+1} \leqq L(g_{k+1} - g_k),$$

where L is a constant, and such that

$$(11) \quad 0 < \zeta < [(g_{k+1} - g_k)/b_{g_k}] < M, \qquad 0 < \zeta < [(g_{k+1} - g_k)/B_{g_k}] < M.$$

THEOREM II. *A necessary and sufficient condition that the series* (2) *which is assumed quasi-monotone in the mean, converge is that* $\sum (g_{k+1} - g_k) a_{g_k}$ *converge.*

Proof. If series (2) converges it converges absolutely.
Next by Lemma 1 and the fact that $b_{n+1} \geqq b_n$ we see that

$$\begin{aligned} 2^{\frac{1}{2}} | a_{g_k+1} + a_{g_k+2} + \cdots + a_{g_{k+1}} | &\geqq t_{g_k+1} + t_{g_k+2} + \cdots + t_{g_{k+1}} \\ &\geqq t_{g_{k+1}}[1 + (1-1/b_{g_k}) + (1-1/b_{g_k})^2 + \cdots + (1-1/b_{g_k})^{g_{k+1}-g_k-2}] \\ &= b_{g_k} t_{g_{k+1}}[1 - \{(1-1/b_{g_k})^{b_{g_k}}\}^{(g_{k+1}-g_k-1)/b_{g_k}}] \\ &\geqq c b_{g_k} | a_{g_{k+1}} |, \qquad c \neq 0. \end{aligned}$$

---

[2] $[n/2]$ denotes the largest integer not greater than $n/2$.

Similarly,

$$
\begin{aligned}
| a_{g_k} + t_{g_{k+1}} + \cdots + a_{g_{k+1}-1} | &\leqq t_{g_k} + t_{g_{k+1}} + \cdots + t_{g_{k+1}-1} \\
&\leqq t_{g_k}[1 + (1 + 1/B_{g_k}) + \cdots + (1 + 1/B_{g_k}) \cdots (1 + 1/B_{g_{k+1}-1})] \\
&\leqq - t_{g_k} B_{g_k}[1 - \{(1 + 1/B_{g_k})^{B_{gk}}\}^{(g_{k+1}-g_k+1)/B_{gk}}] \\
&\leqq C t_{g_k} B_{g_k} \leqq 2^{\frac{1}{2}} C \mid a_{g_k} \mid B_{g_k}.
\end{aligned}
$$

The theorem follows by virtue of these relations and of (10) and (11).

We next proceed to a generalization of the Cauchy integral test. Let $b(\alpha)$ be a real monotonic increasing function defined when $\alpha \geqq 0$ and always greater than 1.

We assume a function $a(\alpha) = x(\alpha) + iy(\alpha)$ and let $t(\alpha) = x(\alpha) + y(\alpha)$. We call $a(\alpha)$ quasi-monotone in the mean if $x(\alpha) \geqq 0$, $y(\alpha) \geqq 0$ and

$$
(12) \qquad t(\alpha + \beta) + t(\alpha + \beta + 1) \geqq 2t(\alpha + 2)[1 - 1/b(\alpha + 1)]
$$

so long as $\alpha \geqq 0$ and $0 \leqq \beta \leqq 1$. We write (12) for brevity

$$
t(\alpha + \beta) + t(\alpha + \beta + 1) \geqq P(\alpha) t(\alpha + 2).
$$

All functions are assumed integrable in the Riemann sense.

THEOREM III. *A necessary and sufficient condition that* $\sum_{a=1}^{\infty} a(\alpha)$ *converge is that* $\int_1^{\infty} a(\alpha) d\alpha$ *converge.*

Proof of Theorem III. We have

$$
\begin{aligned}
(13) \qquad \sum_{a=1}^{n} \int_0^1 [t(\alpha + \beta) + t(\alpha + \beta + 1)] d\beta &= \int_1^{n+1} [t(\alpha) + t(\alpha + 1)] d\alpha \\
&\geqq \sum_{u-1}^{n} P(\alpha) t(\alpha + 2) \geqq P(1) \sum_{a=1}^{n} t(\alpha + 2).
\end{aligned}
$$

Next let $\alpha + \beta = \sigma$ and $1 - \beta = \tau$. Then

$$
t(\sigma) + t(\sigma + 1) \geqq P(\sigma + \tau - 1) t(\sigma + \tau + 1).
$$

Hence

$$
(14) \qquad \int_2^{n+1} t(\alpha) d\alpha \leqq \sum_{\sigma=1}^{n} [t(\sigma) + t(\sigma + 1)]/P(0).
$$

The theorem follows from (13) and (14) and the relation,

$$
t(\alpha) \geqq | a(\alpha) | \geqq 2^{-\frac{1}{2}} t(\alpha).
$$

THE UNIVERSITY OF GEORGIA.

SOME THEOREMS ON THE DIMENSION OF FIBRE SPACES.*

By S. D. Liao.

The aim of this paper is to establish some theorems which are concerned with the relationship between the dimensions of a fibre space, its base space, and its fibres. The notion of a fibre space is understood in the sense of Hurewicz-Steenrod [1]. Following their paper we shall make consistent use of the notations:

$X =$ fibre space, σ being its metric;

$B =$ base space, ρ being its metric;

$\pi =$ projection of X on B;

$\phi(x, b)$ is the slicing function, so that $\phi(x, b)$ is defined for $x \in X$, $b \in B$, with $\rho(\pi(x), b) < \epsilon_0$.

We shall always suppose that both X and B are separable.

Let T be a separable metric space and S a subset of T. The least dimension of the open subsets W of T with $W \supset S$ will be denoted by $\dim(S, T)$. With this notation we shall state the theorems which we intend to prove in this paper as follows:

THEOREM 1. *Let a separable metric space X be a fibre space over a polyhedron B with compact fibres. Then for any point b of B,*

I) $$\mathit{dim}\,(b, B) + \dim \pi^{-1}(b) \leqq \dim(\pi^{-1}(b), X).$$

THEOREM 2. *Let a finite dimensional compactum X be a fibre space over a polyhedron B. The subset B^* of B which consists of all points b of B for which* $\dim(b, B) + \dim \pi^{-1}(b) = \dim(\pi^{-1}(b), X)$ *holds, contains an open dense subset of B.*

THEOREM 3. *Under the same hypotheses as Theorem 2, let k denote the greatest dimension of the fibres of X. If B has homogeneous dimension, then $n + k = m$, where $m = \dim X$, $n = \dim B$.*

THEOREM 4. *Let a compactum X of dimension $m > 0$ be a fibre space*

---

* Received January 17, 1948.

over a polyhedron B *of dimension* n. *If* X *possesses the following property*:
J) *Any compact neighborhood retract* T *of* X *of dimension* $< m$ *which disconnects* X *has non-vanishing* $(m-1)$*-dimensional Čech homology group* $H^{m-1}(T, \mathfrak{R})$,[1]

then: (i) *all the fibres of* X *have the same dimension* $m-n$, *and* (ii) B *is a strongly connected homogeneous* n*-dimensional polyhedron.*

THEOREM 5. *An irreducible closed polyhedron has the property* J *of Theorem* 4. *Hence the conclusions of Theorem* 4 *hold, if it is a fibre space over a polyhedron.*

Theorems 1, 2 and 3 give the relationship between the dimensions of X, B and $\pi^{-1}(b)$, while Theorems 4 and 5 give sufficient conditions that the fibres have the same dimension.

1. Proof of Theorem 1.

LEMMA 1. *Let a metric space* X *be a fibre space over a metric space* B *with compact fibres. Let* b *be an arbitrary point of* B *and* ζ *an arbitrary positive real number. There is a positive real number* $\xi < \epsilon_0$ *such that* $\sigma(\phi(x, u), x) < \zeta$ *for any point* (x, u) *of* $X \times B$ *with* $x \,\varepsilon\, \pi^{-1}(b)$ *and* $\rho(b, u) < \xi$.

Proof. Suppose the contrary. There is a sequence $\{\xi_n\}$ of positive real numbers $\xi_n < \epsilon_0$, which converges to 0, and there is a sequence $\{(x_n, u_n)\}$ of points (x_n, u_n) of $X \times B$ with $x_n \,\varepsilon\, \pi^{-1}(b)$ and $\rho(b, u_n) < \xi_n$ such that $\sigma(\phi(x_n, u_n), x_n) \geqq \zeta$, $n = 1, 2, 3, \cdots$. Since $\pi^{-1}(b)$ is a compactum, we can, by changing to a subsequence if necessary, assume that the sequence $\{x_n\}$ in $\pi^{-1}(b)$ converges to a point x_0 of $\pi^{-1}(b)$. Then we have $\lim_{n\to\infty} \phi(x_n, u_n) = \phi(x_0, b)$ and $\lim \sigma(\phi(x_n, u_n), x_n) = 0$ which contradicts the inequalities $\sigma(\phi(x_n, u_n), x_n) \geqq \zeta$, $n = 1, 2, 3, \cdots$. This proves Lemma 1.

Proof of Theorem 1. Clearly there is an open subset U_0 of X with $U_0 \supset \pi^{-1}(b)$ and $\dim U_0 = \dim (\pi^{-1}(b), X)$. There is a positive real number ζ such that U_0 contains the ζ-neighborhood of $\pi^{-1}(b)$. Let $\xi < \epsilon_0$ be the number corresponding to ζ and b according to Lemma 1. We decompose the polyhedron B into a (locally finite) complex K of mesh $< \xi$ (which means that every cell of K is of diameter $< \xi$). Let M be the closed star of b with respect to K. It follows that $\phi(x, u) \,\varepsilon\, U_0$ for all $x \,\varepsilon\, \pi^{-1}(b)$, $u \,\varepsilon\, M$, or by setting $P = \pi^{-1}(b) \times M$, that $\phi(P) \subset U_0$.

---

[1] $\mathfrak{R}$ is the group of the real numbers modulo 1.

Since M is a polyhedron and $\pi^{-1}(b)$ is a compactum, we have[2] $\dim P = \dim \pi^{-1}(b) + \dim M$. As $\dim M = \dim (b, B)$, to prove the inequality I, it is sufficient to prove that $\dim P \leqq \dim U_0$.

Let ϵ be an arbitrary real number. Clearly P is a compactum and the partial mapping $\phi \mid P$ of P into X is uniformly continuous. Hence there exists a number $\delta > 0$ such that $\sigma(\phi(x, u_1), \phi(x, u_2)) < \epsilon/2$ whenever $\rho(u_1, u_2) < \delta$, where (x, u_1), $(x, u_2) \varepsilon P$. We choose an η, $0 < \eta \leqq 1$ so that $\eta\rho(u, b) < \delta$ holds for any $u \varepsilon M$. Then we construct the continuous mapping $g(x, u) = \phi(x, \eta u + (1 - \eta) b)$, $(x, u) \varepsilon P$, and assert that g is an ϵ-mapping of P into U_0. In fact, let $g(x_1, u_1) = g(x_2, u_2)$. We get $\eta u_1 + (1 - \eta) b = \eta u_2 + (1 - \eta) b$, and hence $u_1 = u_2$. Thus,

$$+\sqrt{\sigma(x_1, x_2)^2 + \rho(u_1, u_2)^2} = \sigma(x_1, x_2) \leqq \sigma(x_1, g(x_1, u_1)) + \sigma(x_2, g(x_2, u_2))$$
$$= \sigma(\phi(x_1, b), \phi(x_1, \eta u_1 + (1 - \eta) b)) + \sigma(\phi(x_2, b), \phi(x_2, \eta u_2 + (1 - \eta) b))$$
$$< \epsilon.$$

It follows that for any given $\epsilon > 0$ there is an ϵ-mapping of P into U_0. Therefore[3] $\dim P \leqq \dim U_0$ and Theorem 1 is proved.

Remark 1. Theorem 1 does not remain true if the polyhedron B is replaced by a general separable metric space. In fact, Pontrjagin[4] has defined compacta B_1 and B_2 of dimension 2, whose product space $B_1 \times B_2$ is of dimension 3. Considering $B_1 \times B_2$ as a fibre space over B_1, the inequality I no longer holds.

Remark 2. Without the assumption that the fibres are compact, we can establish the following facts: 1) $\dim (b, B) \leqq \dim (\pi^{-1}(b), X)$ for any $b \varepsilon B$; 2) $\dim B \leqq \dim X$. To prove 1), let U_0 be an open subset of X with $U_0 \supset \pi^{-1}(b)$ and $\dim U_0 = \dim (\pi^{-1}(b), X)$, and let $x_0 \varepsilon \pi^{-1}(b)$. A topological mapping g in X of the ϵ_0-neighbourhood $V(b, \epsilon_0)$ of b in B can be defined by $g(u) = \phi(x_0, u)$, $u \varepsilon V(b, \epsilon_0)$. Let $U'_0 = U_0 \cap g(V(b, \epsilon_0))$. Then $\dim U'_0 \leqq \dim U_0$. Clearly $g^{-1}(U'_0)$ is an open subset of B, containing b, so that $\dim (b, B) \leqq \dim g^{-1}(U'_0)$. This proves 1). 2) is also easily established.

---

[2] For a non-empty compactum Y and a non-empty polyhedron Z, the relation $\dim (Y \times Z) = \dim Y + \dim Z$ is easily deduced from the well known fact [2, p. 34] that the dimension of the product space of a compactum and a 1-dimensional separable metric space is the sum of the two factor spaces.

[3] This follows from an argument given in [3], pp. 364-365.

[4] L. Pontrjagin, "Sur une hypothèse fondamentale de la théorie de la dimension," *Comptes Rendus*, vol. 190 (1930). pp. 1105-1107.

2. Proofs of Theorems 2 and 3.

LEMMA 2. *Let a metric space X be a fibre space over a metric space B with compact fibres and n, an integer $\geqq 0$. The set B_n of points u of B with $\dim \pi^{-1}(u) \geqq n$ is open in B.*

Proof. It suffices to prove that the set B'_n of points u of B with $\dim \pi^{-1}(u) < n$ is closed in B. Let $b \,\varepsilon\, B$ be a limit point of B'_n and ζ, an arbitrary positive real number. Let $\xi < \epsilon_0$ be the number corresponding to b and $\zeta/2$ according to Lemma 1. There is a point u_0 of B'_n with $\rho(b, u_0) < \xi$. Hence $\sigma(\phi(x, u_0), x) < \zeta/2$ for any $x \,\varepsilon\, \pi^{-1}(b)$. We define the continuous mapping g of $\pi^{-1}(b)$ into $\pi^{-1}(u_0)$ by $g(x) = \phi(x, u_0)$, $x \,\varepsilon\, \pi^{-1}(b)$. If x_1 and x_2 are points of $\pi^{-1}(b)$ with $g(x_1) = g(x_2)$, then $\sigma(x_1, x_2) \leqq \sigma(\phi(x_1, u_0), x_1) + \sigma(\phi(x_2, u_0), x_2)$. This shows that g is a ζ-mapping in $\pi^{-1}(u_0)$ of the compactum $\pi^{-1}(b)$. Since $u_0 \,\varepsilon\, B'_n$, $\dim \pi^{-1}(u_0) < n$. It follows that[3] $\dim \pi^{-1}(b) < n$ and hence $b \,\varepsilon\, B'_n$. This proves Lemma 2.

Proof of Theorem 2. X being a compactum, the projection π is a closed mapping of X on B. Let C be any subset of B and let $D = \pi^{-1}(C)$. The partial mapping $\pi \mid D$ will also be a closed mapping of D on C.

We take a cellular decomposition K of the polyhedron B. Let B_0 be the set of all points of B, each of which is interior to a ground cell of K, and let $B' = B_0 \cap B^*$. To prove Theorem 2, it is sufficient to prove that B' is open and dense in B.

Let V be any non-empty open subset of B. B_0 being open and dense in B, $V' = V \cap B_0$ is a non-empty open subset of B. It follows that there is a point b of B such that

$$(1) \qquad b \,\varepsilon\, V', \text{ and } \dim \pi^{-1}(b) \geqq \dim \pi^{-1}(u) \text{ for any } u \,\varepsilon\, V'.$$

Let V_0 be a neighbourhood of b in B such that

$$(2) \qquad V_0 \subset V', \text{ and } \dim V_0 = \dim (b, B).$$

Let $U_0 = \pi^{-1}(V_0)$. Then U_0 is an open subset of X containing $\pi^{-1}(b)$, and we have

$$(3) \qquad \dim U_0 \geqq \dim (\pi^{-1}(b), X).$$

Since the partial mapping $\pi \mid U_0$ is a closed mapping of U_0 on V_0, we conclude by making use of a well-known theorem [2, p. 92], that

$$\dim V_0 + \dim \pi^{-1}(b') \geqq \dim U_0 \text{ for a certain } b' \varepsilon V_0. \tag{4}$$

Combining the inequalities (1)-(4), we get $\dim(b, B) + \dim \pi^{-1}(b) \geqq \dim(\pi^{-1}(b), X)$. On the other hand, by Theorem 1, we have $\dim(b, B) + \dim \pi^{-1}(b) \leqq \dim(\pi^{-1}(b), X)$. It follows that $b \varepsilon B^*$, whence $b \varepsilon B'$. Thus any non-empty open subset of B contains a point of B'. B' is therefore dense in B.

To prove that B' is open in B, we take an arbitrary point $b \varepsilon B'$. Then

$$\dim(b, B) + \dim \pi^{-1}(b) = \dim(\pi^{-1}(b), X), \tag{5}$$

and b belongs to the interior G of a ground cell of K. Evidently, G is an open subset of B, contained in B_0, and

$$\dim(u, B) = \dim(b, B) \text{ for any } u \varepsilon G. \tag{6}$$

By Lemma 2, we have a neighborhood V of b in B such that

$$V \subset G, \text{ and } \dim \pi^{-1}(u) \geqq \dim \pi^{-1}(b) \text{ for any } u \varepsilon V. \tag{7}$$

Clearly, there is an open subset U of X such that

$$\pi^{-1}(b) \subset U \subset \pi^{-1}(V), \text{ and } \dim(\pi^{-1}(b), X) = \dim U. \tag{8}$$

Since π is a closed mapping of X on B, we easily establish that the set $V' = B - \pi(X - U)$ is an open subset of B, containing b, and that

$$V' \subset V,\ \pi^{-1}(V') \subset U. \tag{9}$$

Now let b' be any point of V'. Then $\pi^{-1}(b') \subset \pi^{-1}(V') \subset U$ and $\dim U \geqq \dim(\pi^{-1}(b'), X)$. From (5)-(9), it follows that $\dim(b', B) + \dim \pi^{-1}(b') \geqq \dim(\pi^{-1}(b'), X)$. But by Theorem 1, we have $\dim(b', B) + \dim \pi^{-1}(b') \leqq \dim(\pi^{-1}(b'), X)$. Therefore $b' \varepsilon B^*$ and hence $V' \subset B^*$. It follows that $V' \subset B^* \cap V \subset B^* \cap G \subset B^* \cap B_0 = B'$. Thus we have shown that any point b of B' has a neighbourhood V' in B, contained in B'. B' is therefore open in B. Theorem 2 is thus proved.

Proof of Theorem 3. Clearly, $\dim \pi^{-1}(u) \leqq k$ for any $u \varepsilon B$ and there is a point b of B with $\dim \pi^{-1}(b) = k$. Since B is of homogeneous dimension n, we have $\dim(b, B) = n$, and since π is a closed mapping of X on B, there is a point b' of B such that $n + \dim \pi^{-1}(b') \geqq m$. Thus, $\dim(b, B)$

$+ \dim \pi^{-1}(b) = n + k \geqq n + \dim \pi^{-1}(b') \geqq m \geqq \dim (\pi^{-1}(b), X)$. But, by Theorem 1, $\dim (b, B) + \dim \pi^{-1}(b) \leqq \dim (\pi^{-1}(b), X)$. Combining these two inequalities, we see that $n + k = m$.

COROLLARY. *Under the same hypotheses as Theorem 2, let K be a cellular decomposition of B and B' a closed subpolyhedron of B built out of some proper sides of ground cells of K. Then $\dim \pi^{-1}(B') < m$, where $m = \dim X$.*

Proof. Let[5] $c_1, c_2, \cdots, c_l$ be the cells of K which compose the subpolyhedron B' of B. Then, $B' = \bigcup_{1 \leqq i \leqq l} c_i$, and $\pi^{-1}(B') = \bigcup_{1 \leqq i \leqq l} \pi^{-1}(c_i)$, each $\pi^{-1}(c_i)$ being a closed subset of X. Suppose that $\dim \pi^{-1}(B') \geqq m$. Then, there is at least one c_i, say c_1, such that $\dim \pi^{-1}(c_1) \geqq m$. Clearly $\pi^{-1}(c_1)$ is a compactum of dimension m and is a fibre space over c_1 with projection π, and c_1 is a polyhedron of homogeneous dimension. By Theorem 3, we have $\dim c_1 + k_1 = m$ where k_1 denotes the greatest dimension of the fibres of the fibre space $\pi^{-1}(c_1)$. By hypothesis, there is a ground cell c of K with c_1 as a proper side. We have then $\dim c_1 < \dim c$ and $\dim \pi^{-1}(c_1) \leqq \dim \pi^{-1}(c)$ $(\geqq m)$. Just as we have done for the fibre space $\pi^{-1}(c_1)$, we have for the fibre space $\pi^{-1}(c)$ over c, $\dim c + k = m$ where k denotes the greatest dimension of the fibres of $\pi^{-1}(c)$. Clearly every fibre of the fibre space $\pi^{-1}(c_1)$ coincides with a fibre of the fibre space $\pi^{-1}(c)$, and hence $k_1 \leqq k$. Thus, from $\dim c_1 < \dim c$, we deduce $\dim c_1 + k_1 < \dim c + k$ which contradicts the relation $\dim c_1 + k_1 = m = \dim c + k$. This proves $\dim \pi^{-1}(B') < m$.

3. Proofs of Theorems 4 and 5.

These theorems are concerned with the question whether the fibres of a fibre space have the same dimension. This is in general not the case, as was shown by an example given by Hurewicz and Steenrod [1, p. 64]. Some sufficient conditions, in the form of Theorems 4 and 5, are, however, found to ensure that the fibres have the same dimension.

We shall precede the proofs of the theorems by a number of Lemmas.

LEMMA 3. *Let A_1 and A_2 be compacta of dimensions $\leqq n_1$ and $\leqq n_2$ respectively with $n_1 \geqq 1$. The Čech homology group $H^{n_1+n_2}(A_1 \times A_2, \mathfrak{R})$[1] vanishes if the Čech homology group $H^{n_1}(A_1, \mathfrak{R})$ vanishes.*

---

[5] We use the same symbol c to stand for the cell of K as well as the closed convex set of the polyhedron.

This Lemma is essentially a generalization of Künneth's theorem to Čech homology theory. It can be established by a fairly standard procedure in Čech theory. We shall omit its proof here.

LEMMA 4. *Let X be a fibre space over a polyhedron B with cellular decomposition K, and F a subset of B which is contained in $V(b,\epsilon_0)\cap M$ where $V(b,\epsilon_0)$ is the ϵ_0-neighbourhood of b in B and M is the closed star of b with respect to K. Then the subset $\pi^{-1}(F)$ of X and the product space $P=\pi^{-1}(b)\times F$ have the same homotopy type.*

Proof. We have the continuous mapping τ of $\pi^{-1}(F)$ in P: $\tau(x)=(\phi(x,b),\pi(x))$, $x\,\varepsilon\,\pi^{-1}(F)$, and the continuous mapping θ of P in $\pi^{-1}(F)$: $\theta(x,u)=\phi(x,u)$, $x\,\varepsilon\,\pi^{-1}(b)$, $u\,\varepsilon\,F$. M being the star of b with respect to K, any segment joining b and a point u of F lies in M. We put $f(u,t)=(1-t)u+tb$, $u\,\varepsilon\,F$, $0\leqq t\leqq 1$. Since $F\subset V(b,\epsilon_0)$, the slicing function ϕ is defined for all such points $(x,f(u,t))$ of $X\times B$ with $u\,\varepsilon\,F$ and $0\leqq t\leqq 1$, provided that $\pi(x)$ lies on the segment joining u and b. We then define the homotopy $g(x,t)$ in $\pi^{-1}(F)$ such that $g(x,0)=x$ and $g(x,1)=\theta\tau(x)$ for all $x\,\varepsilon\,\pi^{-1}(F)$ as follows:

$$g(x,t)=\phi(\phi(x,f(\pi(x),t)),\pi(x)),\qquad x\,\varepsilon\,\pi^{-1}(F),\qquad 0\leqq t\leqq 1.$$

Similarly, a homotopy $h(x,u,t)$ is defined in P such that $h(x,u,0)=(x,u)$, $h(x,u,1)=\tau\theta(x,u)$ for all $(x,u)\,\varepsilon\,P$ as follows:

$$h(x,u,t)=(\phi(\phi(x,f(u,1-t)),b),u),\qquad (x,u)\,\varepsilon\,P,\qquad 0\leqq t\leqq 1.$$

These homotopies prove Lemma 4.

LEMMA 5. *Let X be a fibre space over a metric space B, and C a compact neighbourhood retract of B. Then $\pi^{-1}(C)$ is a neighbourhood retract of X.*

Proof. C being a neighbourhood retract of B, there is an open subset V of B with $C\subset V$ and there is a continuous mapping f of V in C with $f(u)=u$ for all $u\,\varepsilon\,C$. Making use of the fact that C is compact, it follows from a well known argument that there is a positive real number δ such that $\rho(f(u),u)<\epsilon_0$ for all $u\,\varepsilon\,V\cap V(C,\delta)$ where $V(C,\delta)$ is the δ-neighbourhood of C in B. Put $V'=V\cap V(C,\delta)$. $\pi^{-1}(V')$ is obviously an open subset of X, containing $\pi^{-1}(C)$. We define a neighbourhood retraction g of $\pi^{-1}(C)$ in X by $g(x)=\phi(x,f(\pi(x)))$, $x\,\varepsilon\,\pi^{-1}(V')$. Therefore $\pi^{-1}(C)$ is a neighbourhood retract of X.

Proof of Theorem 4. The empty set E is obviously a compact neighbourhood retract of X of dimension $< m$ with $H^{m-1}(E, \mathfrak{R}) = 0$. Hence J) implies that X is connected. It follows that the polyhedron B is also connected and arcwise connected. All fibres of X have therefore the same homotopy type.

If $n = 0$, Theorem 4 is obvious. We therefore assume $n > 0$. The polyhedron B can be decomposed into a complex K of mesh $< \epsilon_0$ so that for any $b \varepsilon B$, the closed star M_b of b with respect to K is a proper subset of B. Let b be an arbitrary point of B. Denote by N_b the boundary of M_b. Clearly N_b disconnects B, and is a compact neighbourhood retract of B. Thus, $\pi^{-1}(N_b)$ disconnects X, and is a compact neighbourhood retract of X by Lemma 5. Making use of the Corollary to Theorem 3, we have $\dim \pi^{-1}(N_b) < m$. It follows from J) that $H^{m-1}(\pi^{-1}(N_b), \mathfrak{R}) \neq 0$. Consider the product space $P_b = \pi^{-1}(b) \times N_b$. By Lemma 4, P_b and $\pi^{-1}(N_b)$ have the same homotopy type, whence $H^{m-1}(P_b, \mathfrak{R}) \approx H^{m-1}(\pi^{-1}(N_b), \mathfrak{R})$ and we have

$$(10) \qquad H^{m-1}(P_b, \mathfrak{R}) \neq 0, \text{ for any } b \varepsilon B.$$

Clearly,[2] $\dim P_b = \dim \pi^{-1}(b) + \dim N_b$, and $\dim N_b = \dim (b, B) - 1$. From Theorem 1, $\dim \pi^{-1}(b) + \dim (b, B) \leqq \dim (\pi^{-1}(b), X) \leqq \dim X = m$. On the other hand, (10) gives $\dim P_b \geqq m - 1$. It follows that

$$(11) \qquad \dim \pi^{-1}(b) + \dim N_b = m - 1, \text{ for any } b \varepsilon B.$$

Let k denote the greatest dimension of the fibres of X. We say that

$$(12) \qquad H^k(\pi^{-1}(b), \mathfrak{R}) \neq 0, \text{ for any } b \varepsilon B.$$

(12) is obvious when $k = 0$, for $\pi^{-1}(b)$ is non-empty. Consider the case $k > 0$. There is a point b_0 of B with $\dim \pi^{-1}(b_0) = k$. If $H^k(\pi^{-1}(b_0), \mathfrak{R}) = 0$, then we shall obtain by making use Lemma 3 and by (11), that $H^{m-1}(P_{b_0}, \mathfrak{R}) = 0$, contradictory to (10). Thus $H^k(\pi^{-1}(b_0), \mathfrak{R}) \neq 0$. Since, for any $b \varepsilon B$, $\pi^{-1}(b)$ and $\pi^{-1}(b_0)$ have the same homotopy type, we have (12). It follows immediately from (12) that all fibres of X have the same dimension k. By choosing b to be an interior point of a ground cell of K of dimension n, we get $k = m - n$. This proves (i).

To prove (ii), we first notice that $\dim (b, B) = n$ for any $b \varepsilon B$, i. e., B is a polyhedron of homogeneous dimension n. Let B be covered by two closed proper subpolyhedra L_1 and L_2 built out of cells of K such that no ground cells of K are contained in both L_1 and L_2, and let $B_0 = L_1 \cap L_2$. Then $\pi^{-1}(B_0)$ disconnects X. By the Corollary to Theorem 3, we have

$\dim \pi^{-1}(B_0) \leqq m-1$, and by similar considerations as before, we can establish

$$\dim \pi^{-1}(B_0) = m - 1. \tag{13}$$

Putting $X_0 = \pi^{-1}(B_0)$, the partial mapping $\pi \mid X_0$ is then a closed mapping of X_0 on B_0. Thus, there is a point b_0 of B_0 such that $\dim B_0 + \dim \pi^{-1}(b_0) \geqq \dim \pi^{-1}(B_0)$, and hence from (i) and (13), we obtain $\dim B_0 \geqq n-1$. This proves (ii).

COROLLARY. *In Theorem 4, let* $n > 0$. *We also have for any* $b \varepsilon B$ *the consequences*: (iii) $H^{m-n}(\pi^{-1}(b), \mathfrak{R}) \neq 0$; (iv) *the* $(n-1)$*-dimensional homology group of* B *at* b [4, p. 121] *with coefficients in* $\mathfrak{R}$ *does not vanish.*

(iii) follows from (12) and $k = m - n$. (iv) is obvious for $n = 1$. For $n > 1$, we have $H^{n-1}(N_b, \mathfrak{R}) \neq 0$; otherwise, by (11) and by making use of Lemma 3, we shall get that $H^{m-1}(\pi^{-1}(b) \times N_b, \mathfrak{R}) = 0$, contradicting (10).

Proof of Theorem 5. Let $\dim X = m$. Let T be an arbitrary compact neighbourhood retract of X of dimension $< m$ which disconnects X and p_1, p_2 two points of $X - T$, which belong to different components. There exists a sufficiently fine simplicial decomposition K of X such that the following conditions are satisfied: 1) If K_1 denotes the subcomplex of K consisting of all simplexes not containing p_1, p_2 and K_0 the subcomplex consisting of all simplexes of K which meet T and their sides, then $\bar{K}_0 \subset \bar{K}_1$; 2) [6] $\bar{K}_0$ can be deformed into T by a deformation $f_t(x) \varepsilon \bar{K}_1$, $0 \leqq t \leqq 1$, $x \varepsilon \bar{K}_0$, such that $f_0(x) = x, f_1(x) \varepsilon T$. K being irreducibly closed,[7] there is an m-cycle z^m of K, with coefficients in the group $\mathfrak{G}_{l^*}$ (= additive group of integers modulo a certain $l^* \geqq 2$) such that $\mid z^m \mid = K$. We denote by W_1, W_2 the components of $X - \bar{K}_0$, which contain p_1, p_2 respectively. Write $z^m = \sum_i v_i \sigma_i$, $v_i \varepsilon \mathfrak{G}_{l^*}$. Let a^m be the chain formed by all the terms $v_i \sigma_i$ of z^m for which $(\operatorname{Int} \sigma_i) \cap W_1 \neq 0$. It is then easy to see that ∂a^m is an $(m-1)$-cycle in K_0.

We say that $\partial a^m \not\sim 0$ in $\bar{K}_1$. In fact, suppose $\partial a^m \sim 0$ in $\bar{K}_1$. By simplicial approximation there is an m-chain b^m in K_1 such that $\partial a^m = \partial b^m$. We see that $c^m = a^m - b^m$ is a non-zero m-cycle such that $\mid c^m \mid$ is a proper subcomplex of K. But this contradicts the fact that K is irreducibly closed.

By the deformation f_t it follows that $f_1(\partial a^m) \not\sim 0$ in T. This means that the singular homology group $\mathfrak{H}^{m-1}(T, \mathfrak{G}_{l*}) \neq 0$. Since T, being a compact neighbourhood retract of a polyhedron, is a locally contractible compactum, its singular and Čech homology groups are isomorphic. Hence the Čech group $H^{m-1}(T, \mathfrak{G}_{l*}) \neq 0$. But the latter is isomorphic to a subgroup of $H^{m-1}(T, \mathfrak{R})$ because $\mathfrak{G}_{l*}$ is isomorphic to a subgroup of $\mathfrak{R}$ and $\dim T < m$. Theorem 5 is thus proved.

---

[6] See [3], p. 343.

[7] For the properties of irreducible closed polyhedra, see [3], pp. 274-287; pp. 329-331.

Remark 3. In a recent paper,[8] Montgomery and Samelson have made the conjecture that there is no compact fibering of the Euclidean space other than a homomorphism. Let us remark here that if an m-dimensional Euclidean space is a fibre space over a polyhedron, then the conclusions (i), (ii) in Theorem 4 and (iii), (iv) in the Corollary to Theorem 4 also hold. These can be established by making use of Theorem 4 with a slight modification. The detailed proofs will be omitted here.

INSTITUTE OF MATHEMATICS,
ACADEMIA SINICA.

REFERENCES.

1. W. Hurewitz and N. Steenrod, "Homotopy relations in fibre spaces," *Proceedings of the National Academy of Sciences*, vol. 27 (1941). pp. 60-64.
2. W. Hurewicz and H. Wallman, *Dimension Theory*, Princeton University Press, 1941.
3. P. Alexandroff and H. Hopf, *Topologie*, Berlin, Springer, 1935.
4. H. Seifert and W. Threlfall, *Lehrbuch der Topologie*, Leipzig, Teubner, 1934.

---

[8] D. Montgomery and H. Samelson, "Fiberings with singularities," *Duke Mathematical Journal*, vol. 13 (1946), pp. 51-56.

SEMILINEAR NORMAL BASIS FOR QUASIFIELDS.*

By TADASI NAKAYAMA.

The writer has given recently a theorem which may well be called the theorem of semilinear normal basis.[1] Namely:[2] Let L be a (finite separable) Galois extension of a field K with Galois group $\mathfrak{G}$, and F a subfield of L (not necessarily containing K) such that $F^{\mathfrak{G}} = F$. Then there exists in L an element, whose $(L:K)$ conjugates with respect to K are linearly independent over F or form a module-basis of L over F according as $(L:F) \geqq$ or $\leqq (L:K)$. The theorem can, for instance, effectively be applied to obtain H. J. Riblet's theorem of differential basis.[3]

The present note extends the theorem to the case of a noncommutative domain with a Galois group in the sense of N. Jacobson.[4] To do so we have first to make some preliminary observations on semilinear group rings over quasifields and their moduli (**1**). Then we obtain the theorem of semilinear normal basis for quasifields in a similar manner as in the commutative case, by combining the preliminary discussion with the theorem on normal basis for quasifields[5] (**2**). The theorem of differential basis can also be transferred to quasifields (**3**).

1. Semilinear group rings over quasifields. Let Ω be a quasifield, and let there be given a finite group $\mathfrak{G}$ of automorphisms of Ω. Let Φ be a subquasifield of Ω such that $\Phi\,\mathfrak{G} = \Phi$, and let $\mathfrak{H}$ be the totality of elements in $\mathfrak{G}$ which leave Φ elementwise fixed. $\mathfrak{H}$ is an invariant subgroup of $\mathfrak{G}$. Semilinear group rings

$$\begin{aligned}(\mathfrak{G}, \Omega) &= G_1\Omega + G_2\Omega + \cdots + G_g\Omega, \\ (\mathfrak{G}, \Phi) &= G_1\Phi + G_2\Phi + \cdots + G_g\Phi\end{aligned} \qquad (g = (\mathfrak{G}:1))$$

---

* Received October 6, 1947.

[1] T. Nakayama, "Halblineare Erweiterung des Satzes der Normalbasis und ihre Anwendung auf die Existenz der derivierten (differentialen) Basis, I," *Proceedings of the Imperial Academy of Tokyo*, vol. 21 (1945); II., *ibid.*, vol. 22 (1946).

[2] This was stated in a slightly stronger form.

[3] H. J. Riblet, "A differential basis for algebraic fields," *American Journal of Mathematics*, vol. 63 (1941).

[4] N. Jacobson, "The fundamental theorem of Galois for quasi-fields," *Annals of Mathematics*, vol. 41 (1940).

[5] T. Nakayama, "Normal basis of a quasi-field," *Proceedings of the Imperial Academy of Tokyo*, vol. 16 (1940).

are defined as usual; $\xi G = G\xi^G$ $(\xi \varepsilon \Omega, \Phi)$. The subring

$$(\mathfrak{H}, \Phi) = H_1\Phi + H_2\Phi + \cdots + H_h\Phi \qquad (h = (\mathfrak{H}:1))$$

is an ordinary group ring of $\mathfrak{H}$ over Φ. Let Z be the center of Φ. Then

$$(\mathfrak{H}, \Phi) = (\mathfrak{H}, Z) \times \Phi \qquad (\text{over } Z).$$

Let $\mathfrak{n}$ be the radical of $(\mathfrak{H}, Z)$, and let

$$(\mathfrak{H}, Z)/\mathfrak{n} = \mathfrak{a}_1 + \mathfrak{a}_2 + \cdots + \mathfrak{a}_k$$

be the decomposition of the semisimple residue ring into simple ideals. We have

$$(\mathfrak{H}, \Phi)/\mathfrak{n}\Phi = \mathfrak{a}_1\Phi + \mathfrak{a}_2\Phi + \cdots + \mathfrak{a}_k\Phi$$

with simple $\mathfrak{a}_i\Phi$, and $\mathfrak{n}\Phi$ forms the radical of $(\mathfrak{H}, \Phi)$.

Now

$$(\mathfrak{G}, \Phi) = G_1(\mathfrak{H}, \Phi) + G_2(\mathfrak{H}, \Phi) + \cdots + G_t(\mathfrak{H}, \Phi)$$

where $\{G_1, G_2, \cdots, G_t\}$ is a representative system of $\mathfrak{G}$ mod. $\mathfrak{H}$. Since $\mathfrak{H}$ and Z are invariant under $\mathfrak{G}$, $(\mathfrak{H}, Z)G = G(\mathfrak{H}, Z)$ whence $\mathfrak{n}G = G\mathfrak{n}$, G being an arbitrary element of $\mathfrak{G}$. Hence

$$\mathfrak{m} = G_1\mathfrak{n}\Phi + G_2\mathfrak{n}\Phi + \cdots + G_t\mathfrak{n}\Phi$$

is a nilpotent (two-sided) ideal in $(\mathfrak{G}, \Phi)$. And

$$(\mathfrak{G}, \Phi)/\mathfrak{m} = G_1(\mathfrak{H}, \Phi)/\mathfrak{n}\Phi + G_2(\mathfrak{H}, \Phi)/\mathfrak{n}\Phi + \cdots + G_t(\mathfrak{H}, \Phi)/\mathfrak{n}\Phi.$$

Furthermore $\mathfrak{a}_i \to G^{-1}\mathfrak{a}_iG$ is a permutation of $\mathfrak{a}_1, \mathfrak{a}_2, \cdots, \mathfrak{a}_k$. Let $\mathfrak{F} = \mathfrak{F}_1$ be the totality of elements G in $\mathfrak{G}$ which map $\mathfrak{a}_1$ on itself, and put $\mathfrak{G} = \mathfrak{F}G_1 + \mathfrak{F}G_2 + \cdots + \mathfrak{F}G_s$. Then $\mathfrak{a}_i = G_i^{-1}\mathfrak{a}_1G_i = \mathfrak{a}_1^{G_i}$ $(i = 1, 2, \cdots, s)$ form a transitivity system under $\mathfrak{G}$. We set $\mathfrak{b} = (\mathfrak{a}_1 + \mathfrak{a}_2 + \cdots + \mathfrak{a}_s)\Phi$ and consider

$$\mathfrak{c} = \mathfrak{G}\mathfrak{b} = G_1\mathfrak{b} + G_2\mathfrak{b} + \cdots + G_t\mathfrak{b},$$

under, however,

ASSUMPTION I_0. *The group $\mathfrak{F}/\mathfrak{H} = \mathfrak{F}_1/\mathfrak{H}$ of automorphisms of Φ is outer, that is, every class in $\mathfrak{F}/\mathfrak{H}$ except the unit class induces an outer automorphism in Φ.*

The system $\mathfrak{c}$ is then a simple ring, as one sees by the well-known argument of crossed product theory. Namely, its $\mathfrak{b}$-double-submoduli $G_\mu\mathfrak{a}_i\Phi$ $(\mu = 1, 2, \cdots, t;\ i = 1, 2, \cdots, s)$ are all irreducible and are mutually non-isomorphic. For, if $G_\mu\mathfrak{a}_i\Phi$ and $G_\nu\mathfrak{a}_j\Phi$ are isomorphic, then obviously $i = j$. Further the multiplication by $G_\mu\mathfrak{a}_iG_\mu^{-1}$ on the left shows that $G_\nu^{-1}G_\mu\mathfrak{a}_iG_\mu^{-1}G_\nu$ must coincide with $\mathfrak{a}_i$ $(=\mathfrak{a}_j)$. Hence $G_\mu^{-1}G_\nu$ induces an automorphism in $\mathfrak{a}_i$, whence in $\mathfrak{a}_i\Phi$. Because of our assumption this automorphism of $\mathfrak{a}_i\Phi$ is outer unless $G_\mu^{-1}G_\nu = 1$. Therefore, as one sees easily, $\mathfrak{a}_i\Phi$-double-moduli $G_\mu\mathfrak{a}_i\Phi$, $G_\nu\mathfrak{a}_i\Phi$ are isomorphic only when $\mu = \nu$. So $\mathfrak{c} = \Sigma G_\mu\mathfrak{a}_i\Phi$ is the (unique) completely reducible so-called ideal decomposition of $\mathfrak{c}$ as $\mathfrak{b}$-double-module, and every non-zero $\mathfrak{b}$-double-submodule contains at least one of $G_\mu\mathfrak{a}_i\Phi$'s. It is now readily seen that if this submodule is a two-sided ideal of $\mathfrak{c}$, then it contains all $G_\mu\mathfrak{a}_i\Phi$, whence it coincides with $\mathfrak{c}$. Thus $\mathfrak{c}$ is simple.

For each transitivity system of $\mathfrak{a}_i$ we obtain an analogous simple ring (under the respective assumption), and $(\mathfrak{G},\Phi)/\mathfrak{G}\mathfrak{n}\Phi = (\mathfrak{G},\Phi)/\mathfrak{m}$ is their direct sum. Hence $\mathfrak{m}$ is indeed the radical of $(\mathfrak{G},\Phi)$.

Thus a $(\mathfrak{G},\Phi)$-right-module is completely reducible, if and only if it is completely reducible as an $(\mathfrak{H},\Phi)$-module; observe that it is annihilated by $\mathfrak{m}$ and $\mathfrak{n}$ simultaneously. Our above analysis shows also that two irreducible right-moduli of $(\mathfrak{G},\Phi)$ are isomorphic, when (and only when) they are isomorphic as $(\mathfrak{H},\Phi)$-moduli; they are isomorphic to irreducible right ideals of the same simple component $\mathfrak{c}$ belonging to the same transitivity system. The same holds for completely reducible moduli too.

As for a general $(\mathfrak{G},\Phi)$-right-module $\mathfrak{S}$, the submodule $\mathfrak{S}\mathfrak{m} = \mathfrak{S}\mathfrak{n}\Phi = \mathfrak{S}\mathfrak{n}$ is the smallest submodule such that the residue module is $(\mathfrak{G},\Phi)$-, as well as $(\mathfrak{H},\Phi)$-, completely reducible. Suppose, now that $\mathfrak{S}$ is isomorphic to $(\mathfrak{G},\Phi)$ as $(\mathfrak{H},\Phi)$-module. $\mathfrak{S}/\mathfrak{S}\mathfrak{m} = \mathfrak{S}/\mathfrak{S}\mathfrak{n}$ is obviously $(\mathfrak{H},\Phi)$-isomorphic to $(\mathfrak{G},\Phi)/\mathfrak{m}$. But then, by the above remark on the completely reducible case it is even $(\mathfrak{G},\Phi)$-isomorphic to $(\mathfrak{G},\Phi)/\mathfrak{m}$. In particular, there exists an element $u \varepsilon \mathfrak{S}$, so that $u(\mathfrak{G},\Phi)\,(\text{mod.}\,\mathfrak{S}\mathfrak{m}) = \mathfrak{S}\,(\text{mod.}\,\mathfrak{S}\mathfrak{m})$. Since $\mathfrak{S}\mathfrak{m}$ is the intersection of all maximal $(\mathfrak{G},\Phi)$-submoduli of $\mathfrak{S}$, then $u(\mathfrak{G},\Phi) = \mathfrak{S}$ too. So $\mathfrak{S}$ is $(\mathfrak{G},\Phi)$-homomorphic to $(\mathfrak{G},\Phi)$; the homomorphism must be an isomorphism, as one sees by observing composition lengths for instance. The same argument holds also when the $(\mathfrak{G},\Phi)$-module $\mathfrak{S}$ is a direct sum of $(\mathfrak{H},\Phi)$-submoduli $(\mathfrak{H},\Phi)$-isomorphic to $(\mathfrak{G},\Phi)$; there comes a certain number of u's instead of the single u.

More generally, it is also valid in the case where $\mathfrak{S}$, a $(\mathfrak{G},\Phi)$-right-module, is a direct sum of a (finite or infinite) number of $(\mathfrak{H},\Phi)$submoduli, each of which is $(\mathfrak{H},\Phi)$-isomorphic to a directly indecomposable right ideal direct component of $(\mathfrak{G},\Phi)$. For this last has a form $e(\mathfrak{G},\Phi)$ with a primitive idempotent element e in $(\mathfrak{G},\Phi)$, and so $\mathfrak{S}/\mathfrak{S}\mathfrak{m} = \mathfrak{S}/\mathfrak{S}\mathfrak{n}$ is a direct sum of $(\mathfrak{H},\Phi)$-submoduli, each $(\mathfrak{H},\Phi)$-isomorphic to a module of the form $e(\mathfrak{G},\Phi)/e(\mathfrak{G},\Phi)\mathfrak{n} = e(\mathfrak{G},\Phi)/e\mathfrak{m}$. Then, again by the above remark on the completely reducible case, it is $(\mathfrak{G},\Phi)$-isomorphic to a direct sum of $(\mathfrak{G},\Phi)$-submoduli, say $\mathfrak{S}_\mu$, which may well be different from the submoduli just observed, each of which is however,

$(\mathfrak{G}, \Phi)$-isomorphic to a module $e(\mathfrak{G}, \Phi)/e\mathfrak{m}$, say $e_\mu(\mathfrak{G}, \Phi)/e_\mu\mathfrak{m}$. Let u_μ be, for each μ, an element in $\mathfrak{s}$ such that $u_\mu (\text{mod. } \mathfrak{s}\mathfrak{m})$ corresponds to $e_\mu (\text{mod. } e_\mu\mathfrak{m})$ in the isomorphism between $\mathfrak{s}_\mu$ and $e_\mu(\mathfrak{G}, \Phi)/e_\mu\mathfrak{m}$. Then these elements u_μ altogether $(\mathfrak{G}, \Phi)$- generate $\mathfrak{s}$, since $\mathfrak{s}\mathfrak{m}$ is contained in all maximal submoduli of $\mathfrak{s}$. Now construct, not in $(\mathfrak{G}, \Phi)$ but abstractly, the (restricted) direct sum $\mathfrak{v} = \Sigma e_\mu(\mathfrak{G}, \Phi)$. $e_\mu \to u_\mu e_\mu$ defines a $(\mathfrak{G}, \Phi)$-homomorphic mapping of $\mathfrak{v}$ onto $\mathfrak{s}$. We want to show that it is really an isomorphism. This is settled easily by the argument of composition length when $\mathfrak{s}$ is $(\mathfrak{G}, \Phi)$-finite. For the general case, however, let $\mathfrak{w}$ be the kernel of the homomorphism. We observe that $\mathfrak{s}$ is as $(\mathfrak{H}, \Phi)$-module a direct sum of submoduli isomorphic to directly indecomposable right ideal direct components of $(\mathfrak{H}, \Phi)$; $\mathfrak{s} = \Sigma \mathfrak{t}_\nu$, $\mathfrak{t}_\nu \cong f_\nu(\mathfrak{H}, \Phi)$ with primitive idempotent elements f_ν in $(\mathfrak{H}, \Phi)$. Let $t_\nu \leftrightarrow f_\nu$ in the isomorphism, and let v_ν be a counter image of t_ν in our $(\mathfrak{G}, \Phi)$-, whence $(\mathfrak{H}, \Phi)$-homomorphic mapping of $\mathfrak{v}$ onto $\mathfrak{s}$. The sum $\Sigma v_\nu f_\nu(\mathfrak{H}, \Phi)$, necessarily direct, is $(\mathfrak{H}, \Phi)$-isomorphic to $\mathfrak{s} = \Sigma \mathfrak{t}_\nu$. Its sum with $\mathfrak{m}$ is direct. So, if $\mathfrak{w} \neq 0$, then $\mathfrak{w} \not\subseteq \mathfrak{v}\mathfrak{m} = \mathfrak{v}\mathfrak{m}$, the intersection of all maximal $(\mathfrak{H}, \Phi)$-submoduli of $\mathfrak{v}$, and therefore necessarily $\mathfrak{w} = 0$, since our mapping induces an isomorphism between $\mathfrak{v}/\mathfrak{v}\mathfrak{m}$ and $\mathfrak{s}/\mathfrak{s}\mathfrak{m}$. Thus we have, under our above assumption,

LEMMA 1. *Let $\mathfrak{s}$, $\mathfrak{r}$ be two $(\mathfrak{G}, \Phi)$-right-moduli, and let $\mathfrak{s}$ be a direct sum of $(\mathfrak{H}, \Phi)$-submoduli $(\mathfrak{H}, \Phi)$-isomorphic to directly indecomposable right ideal direct components of $(\mathfrak{G}, \Phi)$. If then $\mathfrak{r}$ and $\mathfrak{s}$ are $(\mathfrak{H}, \Phi)$-isomorphic, they are $(\mathfrak{G}, \Phi)$-isomorphic too.*

From this follows further

LEMMA 2. *Let $\mathfrak{r}$ and $\mathfrak{s}$ be as in Lemma* 1. *If $\mathfrak{r}$ is a direct summand of $\mathfrak{s}$ as $(\mathfrak{H}, \Phi)$-module, then it is a direct summand in $\mathfrak{s}$ as $(\mathfrak{G}, \Phi)$-module too.*[6]

---

[6] More generally our proof to the lemmas remains valid in the case where $\mathfrak{s}$ (as well as $\mathfrak{r}$) is a direct sum of $(\mathfrak{H}, \Phi)$-submoduli all $(\mathfrak{H}, \Phi)$-isomorphic to directly indecomposable right ideal direct components of a fixed residue ring of $(\mathfrak{G}, \Phi)$. (In this formulation is included the above completely reducible case too.)

In fact, it follows also from the remark that the essential setting in the above proof is the following: Let $\mathfrak{B}$ be a ring with the radical $\mathfrak{N}$ and the decomposition $\mathfrak{B}/\mathfrak{N} = \mathfrak{A}_1 + \mathfrak{A}_2 + \cdots + \mathfrak{A}_k$ of the semisimple residue ring. Let there be given a finite subgroup $\mathfrak{g}$ of the automorphism class group of $\mathfrak{B}$, and consider a crossed product $(\mathfrak{g}, \mathfrak{B})$ (with a factor set). Let it be assumed for every $\mathfrak{A}_i$ that every G $(\neq 1) \varepsilon\, \mathfrak{g}$ mapping $\mathfrak{A}_i$ on itself induces an outer automorphism in $\mathfrak{A}_i$; this amounts to the condition that for every $\mathfrak{g}$-invariant two-sided ideal in the residue ring $\mathfrak{B}/\mathfrak{N}$ the group $\mathfrak{g}$ is outer. Then $\mathfrak{M} = (\mathfrak{g}, \mathfrak{N})$ becomes the radical of $(\mathfrak{g}, \mathfrak{B})$, and Lemmas 1, 2 hold for $(\mathfrak{g}, \mathfrak{B})$-right-moduli, $(\mathfrak{G}, \Phi)$ and $(\mathfrak{H}, \Phi)$ taken place by $(\mathfrak{g}, \mathfrak{B})$ and $\mathfrak{B}$.

It is perhaps of some interest to note that our $(\mathfrak{G}, \Phi)$ is, together with $(\mathfrak{H}, \Phi)$, a Frobenius ring.

Let namely[7]

$$\mathfrak{r}^* = \mathfrak{r}G_1 + \mathfrak{r}G_2 + \cdots + \mathfrak{r}G_t \qquad (t = (\mathfrak{G}:\mathfrak{H}))$$

be the $(\mathfrak{G}, \Phi)$-module induced by $\mathfrak{r}$, in the sense of Frobenius. Similarly, let $\mathfrak{s}^*$ denote the induced module of $\mathfrak{s}$. Evidently $\mathfrak{r}^*$ is a direct summand of the $(\mathfrak{G}, \Phi)$-module $\mathfrak{s}^*$. On the other hand, $\mathfrak{r}^*$ is $(\mathfrak{H}, \Phi)$-isomorphic to $\mathfrak{r}^t$, a direct sum of t isomorphic copies of $\mathfrak{r}$, as one readily verifies by counting the (finite or infinite) numbers of their mutually isomorphic direct indecomposable direct components.[8] Therefore, by Lemma 1, $\mathfrak{r}^*$ and $\mathfrak{r}^t$ are $(\mathfrak{G}, \Phi)$-isomorphic. Similarly $\mathfrak{s}^*$ and $\mathfrak{s}^t$ are $(\mathfrak{G}, \Phi)$-isomorphic. Thus $\mathfrak{r}^t$ is a direct summand of $\mathfrak{s}^t$ as $\mathfrak{G}, \Phi)$-module. Then[9] $\mathfrak{r}$ itself is a direct summand of the $(\mathfrak{G}, \Phi)$-module $\mathfrak{s}$.

2. Semilinear normal basis. Let Ω, Φ, $\mathfrak{G}$, $\mathfrak{H}$ be as above. Namely, Ω be a quasifield, $\mathfrak{G}$ a finite group of its automorphisms, Φ a subquasifield of Ω invariant as a whole under $\mathfrak{G}$, and further, $\mathfrak{H}$ the totality of elements in $\mathfrak{G}$ by which Φ remains elementwise invariant. We make

ASSUMPTION I. *The group $\mathfrak{G}/\mathfrak{H}$ of automorphisms of Φ is outer;* (As a matter of fact, a somewhat weaker assumption in **1** suffices for our purpose.) and

ASSUMPTION II. *The group $\mathfrak{H}$ of automorphisms of Ω is outer.*

Let Ψ be the invariant system of $\mathfrak{H}$ in Ω. Then, because of the Assumption II, $(\Omega:\Psi) = h$ and Ω possesses a right, say, normal basis over Ψ, that is, there exists an element ω in Ω so that $\omega^{H_1}, \omega^{H_2}, \cdots, \omega^{H_h}$ form a linearly independent right basis of Ω over Ψ. Hence Ω and $(\mathfrak{H}, \Psi)$ are isomorphic as $(\mathfrak{H}, \Psi)$-right-moduli; $\Sigma\omega^H\alpha_H \leftrightarrow \Sigma H\alpha_H$. They are so, even more, as $(\mathfrak{H}, \Phi)$-right-moduli. Here $(\mathfrak{H}, \Psi)$ is as an $(\mathfrak{H}, \Phi)$-right-module a direct sum $(\Psi:\Phi)_r$ submoduli isomorphic to $(\mathfrak{H}, \Phi)$. Hence Ω has the same structure as $(\mathfrak{H}, \Phi)$-module. On the other hand, $(\mathfrak{G}, \Phi)$ is a direct sum of $t = (\mathfrak{G}:\mathfrak{H}) = g/h$ submoduli $(\mathfrak{H}, \Phi)$-isomorphic to $(\mathfrak{H}, \Phi)$.

If here $(\mathfrak{G}:\mathfrak{H}) \leqq (\Psi:\Phi)_r$, or, what is the same, $(\mathfrak{G}:1) \leqq (\Omega:\Phi)_r$, then $(\mathfrak{G}, \Phi)$ is a direct summand of Ω as $(\mathfrak{H}, \Phi)$-module. By Lemma 2,

---

[7] Cf. Nakayama, II. l. c. 1, Footnote 4.

[8] Even in $(\mathfrak{H}, \Phi)$-infinite case our structure of $\mathfrak{s}$ implies the uniqueness up to isomorphism of its decomposition into directly indecomposable components. See for instance G. Azumaya, "On generalized semi-primary rings and Krull-Remak-Schmidt theorem," forthcoming in the Japanese Journal of Mathematics.

[9] See footnote 8.

$(\mathfrak{G}, \Phi)$ is then a direct summand of the $(\mathfrak{G}, \Phi)$-module Ω. If however $(\mathfrak{G} : \mathfrak{H}) \geqq (\Psi : \Phi)_r$, then the $(\mathfrak{G}, \Phi)$ module $(\mathfrak{G}, \Phi)$ contains conversely Ω as a direct summand. More precisely, if $(\Psi : \Phi)_r$ is finite and $= q(\mathfrak{G} : \mathfrak{H}) + w$ $(0 \leqq w < (\mathfrak{G} : \mathfrak{H}))$, then the $(\mathfrak{G}, \Phi)$-module Ω is a direct sum of q submoduli isomorphic to $(\mathfrak{G}, \Phi)$ and a module which is a direct summand of $(\mathfrak{G}, \Phi)$. Thus we have

THEOREM. *Let Ω, Φ, $\mathfrak{G}$, $\mathfrak{H}$ be as above, and assume I (or the somewhat weaker I_0 and its analogues for other transitivity systems, which altogether we denote by I_0'), II. If $(\mathfrak{G} : 1) \leqq (\Omega : \Phi)_r$, there exists an element ξ in Ω such that*

$$(*) \qquad \xi^{G_1}, \xi^{G_2}, \cdots, \xi^{G_g} \qquad (\mathfrak{G} = \{G_1, G_2, \cdots, G_g\})$$

are right-, say, linearly independent over Φ. If $(\mathfrak{G} : 1) \geqq (\Omega : \Phi)_r$, however, there exists an element ξ in Ω such that suitable $(\Omega : \Phi)_r$ among (*) form a linearly independent right-basis of Ω over Φ. More precisely, if $(\Omega : \Phi)_r$ is finite and

$$(\Omega : \Phi)_r = q(\mathfrak{G} : 1) + v \qquad (0 \leqq v < (\mathfrak{G} : 1)),$$

then there exist in Ω q elements ξ_i $(i = 1, 2, \cdots, q)$ and an element η so that

$$\xi_i^{G_j} \qquad (i = 1, 2, \cdots, q;\ j = 1, 2, \cdots, g)$$

and suitable v among

$$\eta^{G_j} \qquad (j = 1, 2, \cdots, g)$$

together form a linearly independent right-basis of Ω over Φ.

3. Differential basis. Let Ω be a quasifield, and let there be given in Ω a higher differentiation $\xi \rightarrow \xi^{(\nu)}$ satisfying the iterative law $\binom{\nu}{\mu} \xi^{(\nu)} = \xi^{(\nu)(\nu-\mu)}$, in the sense of F. K. Schmidt.[10] Further, let there be given a group $\mathfrak{G}$ of automorphisms in Ω such that $(\xi^G)^{(\nu)} = (\xi^{(\nu)})^G$ for every $\xi \varepsilon \Omega$, $G \varepsilon \mathfrak{G}$. Denote the subquasifield of (absolute) constants of the differentiation by Φ. We have $\Phi^{\mathfrak{G}} = \Phi$.

LEMMA 3. *If $\xi_1, \xi_2, \cdots, \xi_n$ $(\varepsilon \Omega)$ are right-linearly independent over Φ, then there exist (exactly) n vectors left-independent over Ω among $(\xi_1^{(\nu)}, \xi_2^{(\nu)}, \cdots, \xi_n^{(\nu)})$ $(\nu = 0, 1, 2, \cdots)$.*

---

[10] F. K. Schmidt-H. Hasse, "Noch eine Begründung der Theorie der höheren Differentialquotienten in einem algebraischen Funktionenkörper einer Unbestimmten," *Journal für Mathematik*, vol. 176 (1937).

This is perhaps more or less well-known, but we shall give its proof for the sake of completeness. Let $(\xi_1^{(\nu_i)}, \xi_2^{(\nu_i)}, \cdots, \xi_n^{(\nu_i)})$ with $i = 1, 2, \cdots, m$ be a maximal Ω-left-independent system among the above vectors, and assume $m < n$. We denote the (m, n)-matrix $(\xi_j^{(\nu_i)})_{ij}$ by Ξ. Then, for each ν, $(\xi_1^{(\nu)}, \xi_2^{(\nu)}, \cdots, \xi_n^{(\nu)}) = (\lambda_{\nu 1}, \lambda_{\nu 2}, \cdots, \lambda_{\nu m})\Xi$ with λ's $\varepsilon\, \Omega$. The maximal number s of Ω-right-independent columns in Ξ can not be n;[11] $s < n$. We can assume without loss of generality that the first s columns are independent. Then, for instance, $\xi_n^{(\nu_i)} = \xi_1^{(\nu_i)}\alpha_1 + \xi_2^{(\nu_i)}\alpha_2 + \cdots + \xi_s^{(\nu_i)}\alpha_s$ $(i = 1, 2, \cdots, m)$ with $\alpha_1, \alpha_2, \cdots, \alpha_s \,\varepsilon\, \Omega$. Taking λ-combinations we have

$$(1) \qquad \xi_n^{(\nu)} = \xi_1^{(\nu)}\alpha_1 + \xi_2^{(\nu)}\alpha_2 + \cdots + \xi_s^{(\nu)}\alpha_s$$

for every ν. (μ)-differentiation gives

$$(2) \qquad \xi_n^{(\nu)\,(\mu)} = (\xi_1^{(\nu)}\alpha_1^{(\mu)} + \cdots) \\ + (\xi_1^{(\nu)\,(1)}\alpha_1^{(\mu-1)} + \cdots) + \cdots + (\xi_1^{(\nu)\,(\mu)}\alpha_1 + \cdots).$$

Now, suppose that the 1st, 2nd, $\cdots, \mu - 1$-th derivatives of α_j are all 0. Then only the first and the last sums remain in (2), thus

$$\xi_n^{(\nu)\,(\mu)} - (\xi_1^{(\nu)\,(\mu)}\alpha_1 + \cdots) = (\xi_1^{(\nu)}\alpha_1^{(\mu)} + \cdots).$$

But the left-hand side vanishes, because of the relation (1), ν replaced by $\nu + \mu$, and the iterative law. Hence $\xi_1^{(\nu)}\alpha_1^{(\mu)} + \xi_2^{(\nu)}\alpha_2^{(\mu)} + \cdots + \xi_s^{(\nu)}\alpha_s^{(\mu)} = 0$. Since this is the case for $\nu = 0, 1, 2, \cdots$, in particular for $\nu = \nu_1, \nu_2, \cdots, \nu_m$, necessarily $\alpha_1^{(\mu)} = \alpha_2^{(\mu)} = \cdots = \alpha_s^{(\mu)} = 0$. So we see by induction that all the derivatives $\alpha_j^{(\mu)}$ $(\mu = 1, 2, 3, \cdots)$ vanish; $\alpha_j \,\varepsilon\, \Phi$. Thus ξ_n turns out to be Φ-right-dependent on $\xi_1, \xi_2, \cdots, \xi_s$.

Now we assume I (or I_0') and II, and put $k = \max((\mathfrak{G}:1), (\Omega:\Phi)_r)$. There is, according to our theorem on semilinear normal basis, an element ξ in Ω such that $\xi^{G_1}, \xi^{G_2}, \cdots, \xi^{G_k}$ are right-independent over Φ, where $G_1, G_2, \cdots, G_k$ denote suitably chosen k elements in $\mathfrak{G}$. By the above Lemma 3 there exist $\nu_1, \nu_2, \cdots, \nu_k$ such that

$$(\xi^{(\nu_i)G_1}, \xi^{(\nu_i)G_2}, \cdots, \xi^{(\nu_i)G_k}) \qquad\qquad (i = 1, 2, \cdots, k)$$

are left-independent over Ω. Then $\xi^{(\nu_i)}$ $(i = 1, 2, \cdots, k)$ are left-independent over the invariant system Λ of $\mathfrak{G}$ in Ω. For, any Λ-relation among them would imply the same relation among $\xi^{(\nu_i)G}$ for every G, in particular for each G_j $(j = 1, 2, \cdots, k)$, whence the same relation for the above k vectors. We obtain thus

THEOREM. *Let* Ω, $\mathfrak{G}$, *the differentiation* $\xi \to \xi^{(\nu)}$, Φ (and $\mathfrak{H}$) *be as above,*

---

[11] As a matter of fact $s = m$, as follows from the elementary divisor theorem for instance, but $s < n$ suffices.

and let I (or I_0'), II be assumed. Then there exists an element ξ in Ω whose suitable $k = \max((\mathfrak{G}:1), (\Omega:\Phi)_r)$ derivatives

$$(\dagger) \qquad \xi^{(\nu_1)}, \xi^{(\nu_2)}, \cdots, \xi^{(\nu_k)} \qquad (\nu = 0, 1, 2, \cdots)$$

are left-, say, independent over the invariant system Λ of $\mathfrak{G}$ in Ω.

COROLLARY. *Assume $(\Omega:\Phi)_r \geqq (\mathfrak{G}:1)$, and replace II by a somewhat stronger*

ASSUMPTION II_1. *The group $\mathfrak{G}$ of automorphisms of Ω is outer. Then $k = (\Omega:\Lambda) = g$, and $(\dagger)$, with $k = g$, forms a left-basis of Ω over Λ.*

COROLLARY. *Let (as in II_1) $\mathfrak{G}$ be a finite group of outer automorphisms of a quasifield Ω, and let there be given a differentiation in Ω satisfying the iterative law and commutative with every operation in $\mathfrak{G}$. Suppose further that the subquasifield Φ of constants is contained in the invariant system Λ of $\mathfrak{G}$ in Ω. Then there is an element in Ω such that its suitable $g = (\mathfrak{G}:1)$ derivatives, including the element itself, form a (linearly independent) left-basis of Ω over Λ.*

Remark. If we choose in Lemma 3 the first system, in the sense of lexicographic order of $(\nu_1, \nu_2, \cdots, \nu_k)$, then together with a ν_i there appears in the system every $\nu < \nu_i$ such that $p \nmid \binom{\nu_i}{\nu}$, p being the characteristic of Ω. (In case $p = 0$, this means simply that the system is $(0, 1, \cdots, k-1)$.) This can be seen in the same manner as in the commutative case; see Nakayama, l. c. II. Hilfssatz 2. So the system $(\dagger)$ in the above theorem, or its corollaries, may be taken so as to satisfy this condition.

Throughout in the above we have restricted ourselves to quasifields. However the Galois theory has been extended recently by G. Azumaya and the writer to simple rings, and more generally, to closed irreducible rings.[12] It is also possible, though perhaps not of much interest, to transfer the above to such general cases.

Addendum: [Added in proof] The assumptions I and II for our theorem on semi-linear normal bases may be replaced, as the author has found recently, by the assumption that G forms a group of outer automorphisms in Ω. It is possible to embrace the two cases in one. See forthcoming paper by the author entitled "Galois theory for general rings with minimum condition" and by G. Azumaya entitled "Galois theory for uniserial rings."

NAGOYA IMPERIAL UNIVERSITY.

---

[12] T. Nakayama and G. Azumaya, "On irreducible rings," *Annals of Mathematics*, vol. 48, pp. 949-966.

LEFT ASSOCIATES OF MONIC MATRICES, WITH AN APPLICATION TO UNILATERAL MATRIX EQUATIONS.*[1]

By James H. Bell.

Introduction. Let $\mathfrak{F}[\lambda]$ be the ring of polynomials in the indeterminate λ, over a field $\mathfrak{F}$. This paper will deal with square matrices of order n with elements in $\mathfrak{F}[\lambda]$ unless otherwise specified.

A matrix M with elements in $\mathfrak{F}[\lambda]$ may be written:

$$M = M_k\lambda^k + M_{k-1}\lambda^{k-1} + \cdots + M_0 = \sum_{m=0}^{k} M_m\lambda^m$$

where each M_m $(m = 0, 1, 2, \cdots, k)$ is a matrix with elements in $\mathfrak{F}$. If $M_k \neq 0$, M is said to be of *degree k in λ*. When M_k is non-singular, M is said to be *proper of degree k in λ*. In particular, if M_k is the identity matrix I, the matrix M will be called *monic*[2] *of degree k*.

A matrix T with elements in $\mathfrak{F}[\lambda]$ is *unimodular* if it has an inverse whose elements are also in $\mathfrak{F}[\lambda]$. Two matrices A and B are called *left associates* if there exists a unimodular matrix T such that $TA = B$. *The relationship of left associate is an equals relationship* [1].

The totality of $n \times n$ matrices with elements in $\mathfrak{F}[\lambda]$ may be divided into classes of left associates. Each class is represented by a unique matrix in *canonical triangular form*. This canonical triangular matrix is a matrix having all of the elements below the main diagonal equal to zero. The diagonal elements if they are not zero, are monic polynomials. The elements of each column are reduced modulo the main diagonal element when that element is not zero. If the main diagonal element is zero, then every element of the row, in which it occurs, is zero.[3]

The problem considered is that of determining under what conditions a matrix will be the left associate of a monic matrix of degree k. The special case of this problem, arising when $k = 1$, is of value in the completion of

---

* Received September 11, 1947; revised August 17, 1948.

[1] This problem was considered by the author in his doctoral thesis carried out under M. H. Ingraham at the University of Wisconsin. It was first submitted to the American Journal of Mathematics in June 1941, and in revised form September 11, 1947.

[2] This terminology is adapted from that used by Birkhoff and MacLane in their book: *A Survey of Modern Algebra* (Macmillan 1946), in which a polynomial is termed monic if the coefficient of the term of highest degree is unity.

[3] The Hermite normal form as defined by MacDuffee is the form generally used. However, the canonical triangular form defined here will be used in keeping with that used by Ingraham [2]. It is obtained in the same manner as the Hermite normal form except that the operations are carried out on the columns in reverse order.

the algorithm of M. H. Ingraham [2] for the solution of the unilateral matrix equation $\sum_{m=0}^{r} R_m X^m = 0$ over $\mathcal{J}$.

Necessary and sufficient conditions are obtained in answer to this problem. The canonical triangular form of the given matrix may be readily found [1]. The problem then becomes one of determining whether the canonical triangular matrix $A = \sum_{m=0}^{s} A_m \lambda^m$ is a left associate of a monic matrix of degree k. It is found that if the diagonal elements of A are a_{jj}, $(j = 1, 2, \cdots, n)$ then the degree of $(\prod_{m=1}^{n} a_{mm})$ must equal kn and the degree of $(\prod_{m=1}^{j} a_{mm}) \leq kj$. If these conditions are satisfied, then a necessary and sufficient condition that A be a left associate of a monic matrix of degree k is that the $kn \times sn$ matrix

$$W_k = \left\| \begin{matrix} A_s & A_{s-1} \cdots & A_{s-k+1} \cdots & A_1 \\ 0 & A_s \cdots & A_{s-k+2} \cdots & A_2 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 \quad \cdots & A_s \quad \cdots & A_k \end{matrix} \right\|,$$

with elements in $\mathcal{J}$, be of rank kn.

In the proof of the above conditions, the method for the construction of the unimodular matrix which transforms the matrix A into the monic matrix is given. Details for the construction for the case $k = 1$, are given.

The author has extended the results to the case when $\mathcal{J}$ is a quasi-field. The conditions obtained are the same if, in place of the rank of W_k, the idea of left-row rank, or right-column rank, is substituted. Also, Ingraham's algorithm may be extended to the solution of the unilateral matrix equation over a quasi-field. In this paper, for the sake of brevity, no proofs are presented. The reader is referred to the author's doctoral thesis [3] for greater detail.

1. Left associates of monic matrices.

In the following work T will be used to represent a unimodular matrix. The canonical triangular form of the matrix to be tested will be represented by $A = (a_{ij}) = \sum_{m=0}^{s} A_m \lambda^m$. A monic matrix of degree k will have the form $\lambda^k I + B$, where $B = (b_{ij})$ is of degree less than k. The problem is to determine the conditions under which a unimodular T exists, such that $TA = \lambda^k I + B$.

LEMMA 1. *If two monic matrices are left associates they are identical.*

Suppose there is a matrix U such that

$$U(\lambda^k I + \sum_{m=0}^{k-1} B_m\lambda^m) = \lambda^r I + \sum_{m=0}^{r-1} C_m\lambda^m;$$

then $k \leq r$.

If U is unimodular,

$$\lambda^k I + \sum_{m=0}^{k-1} B_m\lambda^m = U^{-1}(\lambda^r I + \sum_{m=0}^{r-1} C_m\lambda^m)$$

and $k \geq r$.

Therefore, if U is unimodular (that is, if the two monic matrices are left associates) $k = r$, $U = I$ and $B_m = C_m$ $(m = 0, 1, 2, \cdots, k-1)$.

If a matrix is the left associate of a monic matrix, that monic matrix is unique. The canonical triangular form of a matrix is also unique. A monic matrix is non-singular. *Therefore, it follows that the unimodular matrix T which transforms the triangular form into the monic matrix is unique.*

Represent the degree of the element a_{ij} by $d(a_{ij})$. If $a_{jj} \neq 0$, then for $i \neq j$ either $a_{ij} = 0$, or $d(a_{ij}) < d(a_{jj})$.

THEOREM 1. *If A is in canonical triangular form and is the left associate of a monic matrix of degree k, then $d(\prod_{m=1}^{n} a_{mm}) = kn$ and $d(\prod_{m=1}^{j} a_{mm}) \leq kj$.*

If determinants of both sides of the equation $TA = \lambda^k I + B$ are taken, then $|T| \cdot |A| = |\lambda^k I + B|$ where $|\lambda^k I + B|$ is a monic polynomial of degree kn. Since T is unimodular and $|A| = \prod_{m=1}^{n} a_{mm}$ where each a_{mm} is a monic polynomial or zero, it follows that $|T| = 1$, $|A| \neq 0$, and $d(\prod_{m=1}^{n} a_{mm}) = kn$.

The matrix equation $TA = \lambda^k I + B$ may be written as follows:

$$\begin{Vmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{Vmatrix} \begin{Vmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{Vmatrix} = \begin{Vmatrix} \lambda^k I_{11} + B_{11} & B_{12} \\ B_{21} & \lambda^k I_{22} + B_{22} \end{Vmatrix}$$

where T_{11}, A_{11}, I_{11} and B_{11} are $j \times j$ matrices. Since $T_{11}A_{11} = \lambda^k I_{11} + B_{11}$, it follows that $d(\prod_{m=1}^{j} a_{mm}) \leq kj$.

LEMMA 2. *The degree of $T = (t_{ij})$, where $TA = \lambda^k I + B$ and A is in canonical triangular form, is at most equal to k. An element of T is of degree k if and only if it is a main diagonal element and the corresponding main diagonal element of A is unity.*

Assume t_{ij}, the element in the i-th row and j-th column of T, is an element of highest degree. By hypothesis $\sum_{m=1}^{n} t_{im}a_{mj} = \delta_{ij}\lambda^k + b_{ij}$ (δ_{ij} the Kronecker delta). Since $d(b_{ij}) < k$, $d(a_{mj}) < d(a_{jj})$ or $a_{mj} = 0$ (for

$m \neq j$) and $d(t_{ij}) \geq d(t_{im})$; then $t_{ij}a_{jj}$ is the term of highest degree. Therefore,

$$d(t_{ij}) \leq d(t_{ij}a_{jj}) = d(\sum_{m=1}^{n} t_{im}a_{mj}) \leq k.$$

It follows immediately that $d(t_{ij}) = d(t_{ij}a_{jj}) = k$ if and only if $\delta_{ij} \neq 0$ (i. e. $i = j$) and $a_{jj} = 1$.

From this point on, it will be assumed that the elements a_{jj} $(j = 1, 2, \cdots, n)$ satisfy the necessary conditions of Theorem 1. The proof of the second condition given in the introduction will now be carried out.

Since $d(\prod_{m=1}^{n} a_{mm}) = kn$ and $d(\prod_{m=1}^{n-1} a_{mm}) \leq k(n-1)$, then $d(a_{nn}) \geq k$ and $s \geq k$; where s is the degree of A. The elements of A may be written $a_{ij} = \sum_{m=0}^{s} a_{ij,m}\lambda^m$ where $a_{ij,m} = 0$, if $m > d(a_{jj})$ or $i > j$, and $a_{ij,\bar{j}} = \delta_{ij}$ $(\bar{j} = d(a_{jj}))$.[4] The coefficients $a_{ij,m}$ belong to $\mathfrak{F}$.

Consider the i-th row of T. From Lemma 2 $t_{ip} = \delta_{0\bar{i}}\delta_{ip}\lambda^k + \sum_{m=0}^{k-1} t_{ip,m}\lambda^m$ $(p = 1, 2, \cdots, n)$, where $\delta_{0\bar{i}} = 0$ if $d(a_{ii}) = \bar{i} \neq 0$. Then, for a fixed i,

$$\sum_{p=1}^{n} (\delta_{0\bar{i}}\delta_{ip}\lambda^k + \sum_{m=0}^{k-1} t_{ip,m}\lambda^m)(\sum_{m=0}^{s} a_{pj,m}\lambda^m) = \delta_{ij}\lambda^k + b_{ij}.$$

Therefore,

$$\sum_{p=1}^{n} (\sum_{m=0}^{k-1} t_{ip,m}\lambda^m)(\sum_{m=0}^{s} a_{pj,m}\lambda^m) = -\delta_{0\bar{i}}\lambda^k(\sum_{m=0}^{s} a_{ij,m}\lambda^m) + \delta_{ij}\lambda^k + b_{ij}.$$

The coefficients of λ^{s+k} on either side are zero since $\delta_{0\bar{i}}a_{ij,s} = 0$. A system of s equations in kn unknowns $t_{ip,m}$ $(p = 1, 2, \cdots, n;\ m = 0, 1, \cdots, k-1)$ is obtained for a fixed j, by equating coefficients of λ^u $(u = s+k-1,\ s+k-2, \cdots, k)$. They are as follows:

$$(1) \qquad \sum_{m=1}^{k} (\sum_{p=1}^{n} t_{ip,k-m}a_{pj,s+m-v}) = -\delta_{0\bar{i}}a_{ij,s-v} + \delta_{ij}\delta_{sv} \qquad (v = 1, 2, \cdots, s).$$

For each value of v let $j = 1, 2, \cdots, n$. A system of equations sn in number,

$$(2) \qquad \sum_{m=1}^{k} (\sum_{p=1}^{n} t_{ip,k-m}a_{pj,s+m-v}) = -\delta_{0\bar{i}}a_{ij,s-v} + \delta_{ij}\delta_{sv} \qquad (v = 1, 2, \cdots, s;\ j = 1, 2, \cdots, n \text{ for each } v),$$

is obtained.

---

[4] When a_{ij} is written as $\sum_{m=0}^{s} a_{ij,m}\lambda^m$, it will be understood that a_{ij} is written in descending powers of λ.

If the equations are written out in the order given above, the matrix of the coefficients of the kn unknowns $t_{ip,k-m}$ is

$$W_k = \left\| \begin{matrix} A_s & A_{s-1} \cdots & A_{s-k+1} \cdots & A_1 \\ 0 & A_s \cdots & A_{s-k+2} \cdots & A_2 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \quad \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 \quad \cdots & A_s \quad \cdots & A_k \end{matrix} \right\| .$$

Since $d(a_{jj}) = \bar{j}$, $a_{pj,m} = 0$ for $m > \bar{j}$. Therefore, each side of the equation displayed in (1) above, is identically zero for $v = 1, 2, \cdots, s - \bar{j}$. Therefore, the s equations reduce to $\bar{j}$ equations, and the system of sn equations in (2) reduces to $\sum_{j=1}^{n} \bar{j}$ $(= kn)$ non-trivial equations (2′) not displayed here. If these equations have a solution, a matrix T may be constructed so that $TA = \lambda^k I + B$ and since $d(\prod_{m=1}^{n} a_{mm}) = kn$, it follows that T is unimodular. If a unimodular T exists such that $TA = \lambda^k I + B$, the system of equations (2′) has a unique solution. A necessary and sufficient condition that the system have a unique solution is that the rank of the matrix of the system be kn. This, in turn, requires that the rank of W_k be kn.

THEOREM 2. *A necessary and sufficient condition that the canonical triangular matrix* $A = \sum_{m=0}^{s} A_m \lambda^m$, *be the left associate of a monic matrix of degree* k *is that* $d(\prod_{m=1}^{n} a_{mm}) = kn$, $d(\prod_{m=1}^{j} a_{mm}) \leq kj$, *and the rank of the matrix* W_k, *defined above, be equal to* kn.

If the rank of W_k is kn, the matrix T may be obtained by setting up the equations as in (2) and solving for the elements $t_{ip,m}$ $(p = 1, 2, \cdots, n;$ $m = 0, 1, \cdots, k-1)$ for each value of $i = 1, 2, \cdots, n$. Once the coefficients $t_{ip,m}$ have been found, the element t_{ip} of T may be constructed by setting $t_{ip} = \delta_{0\bar{i}} \delta_{ip} \lambda^k + \sum_{m=0}^{k-1} t_{ip,m} \lambda^m$.

The monic matrix is then obtained from the equation $TA = \lambda^k I + B$.

2. Application to the unilateral equation. M. H. Ingraham [2] has shown that the solution of the unilateral matrix equation $\sum_{m=0}^{r} R_m X^m = 0$, over $\mathfrak{F}$, may be carried out in the following manner. The matrix X, with elements in $\mathfrak{F}$, is a solution of the unilateral matrix equation if and only if $\lambda I - X$ is a right divisor of the matrix $R = \sum_{m=0}^{r} R_m \lambda^m$, that is,

$R = M(\lambda I - X)$, where M is a matrix with elements in $\mathfrak{F}[\lambda]$. Both sides of this equality may be multiplied by a unimodular matrix U. Also a unimodular matrix T and its inverse T^{-1} may be inserted without affecting the equality. That is, $UR = UMTT^{-1}(\lambda I - X)$, or $Q = PA$ where $Q = UR$, $P = UMT$ and $A = T^{-1}(\lambda I - K)$. The matrices U and T may be so chosen that Q and A are in canonical triangular form.

The method of solution is to reduce R to its canonical triangular form Q, and then to find the canonical triangular form A of each possible right divisor of Q. Since Q and A are triangular, the matrix P is necessarily triangular. Therefore, $q_{jj} = p_{jj}a_{jj}$ $(j = 1, 2, \cdots, n)$. The elements a_{jj} are selected from the divisors of the main diagonal elements of Q in accordance with Theorem 1, as it is desired to have $TA = \lambda I - X$. The other elements of A are obtained by solving congruences of the form

$$p_{jj}a_{j,j+h} \equiv q_{j,j+h} - \sum_{m=j+1}^{j+h-1} p_{jm}a_{m,j+h} \mod a_{j+h,j+h}. \tag{3}$$

Equation (3) gives a recursive formula by means of which the successive columns of A may possibly be determined. The congruence may have no solution, a unique solution, or many solutions. The criterion for solution of the congruence is that the greatest common divisor of p_{jj} and $a_{j+h,j+h}$ shall divide $q_{j,j+h} - \sum_{m=j+1}^{j+h-1} p_{jm}a_{m,j+h}$. If p_{jj} and $a_{j+h,j+h}$ are relatively prime, there is a unique solution. If the division is impossible at any stage, there is no solution, and there is no right divisor A having the chosen main diagonal elements.

Once a right divisor A is obtained, it is then necessary to determine whether or not it is the left associate of a monic matrix of degree unity. If $A = \sum_{m=0}^{s} A_m\lambda^m$, the matrix W_k becomes W_1 where

$$W_1 = \| A_s \, A_{s-1} \cdots A_1 \| .$$

The necessary and sufficient condition that A be the left associate of a monic matrix of degree 1 *is that W_1 be of rank n.*

If W_1 is of rank n, it is possible to find T such that $TA = \lambda I + B$, where B now has elements in $\mathfrak{F}$. It follows that $X = -B$ is a solution of the unilateral equation.

It is possible to find X without finding T explicitly. When $k = 1$, the equations (2) become

$$\sum_{p=1}^{n} (t_{ip,0}a_{pj,s-v+1}) = -\delta_{0\bar{i}}a_{ij,s-v} + \delta_{ij}\delta_{sv} \qquad (v = 1, 2, \cdots, s;\ j = 1, 2, \cdots, n). \tag{4}$$

Consider i for those values only such that $d(a_{ii}) = \bar{i} > 0$. Designate these values of j by $\alpha, \beta, \gamma, \cdots, \omega$ where $\alpha < \beta < \gamma < \cdots < \omega = n$. Since $a_{pj,m} = 0$ $(m > \bar{j})$, the system of n equations

$$(5) \qquad \sum_{p=1}^{n} t_{ip,0} a_{pj,m} = -\delta_{0\bar{i}} a_{ij,m-1} + \delta_{ij}\delta_{m1} \qquad (m = \bar{j}, \bar{j}-1, \cdots, 1;\ j = \alpha, \beta, \cdots, \omega)$$

is equivalent to the equations in (4). If the matrix $\bar{W}_1$ of the coefficients of the unknowns, $t_{ip,0}$, be written out, it will be seen that it is the matrix which is obtained by these steps:[5]

1. Augment each element of A by proper powers of λ with zero coefficients so that each element in a particular column is a polynomial in descending powers of λ, equal in degree to the main diagonal element.

2. Split up each column into columns, each one involving monomials of the same degree in λ.

3. Delete all columns involving constant terms only, and then set $\lambda = 1$.

A necessary and sufficient condition that there exist a matrix T such that $TA = \lambda I - X$, where $d(\prod_{m=1}^{j} a_{mm}) \leq j$ and $d(\prod_{m=1}^{n} a_{mm}) = n$, is that the rank of $\bar{W}_1$ equal n.

To find X, set up the system of equations $D\bar{W}_1 = K$, where $D = (t_{ip,0})$ and K is constructed as follows: If $a_{ii} = 1$, the i-th row of K is the negative of the i-th row of the matrix obtained from step 2 by deleting the columns involving $a_{jj,\bar{j}}$ $(j = 1, 2, \cdots, n)$ and setting $\lambda = 1$. When $a_{ii} \neq 1$, the i-th row of K has unity in the r-th place, where $r = d(\prod_{m=1}^{i} a_{mm})$, and zero elsewhere. By letting $i = 1, 2, \cdots, n$, each row of K is obtained. If $\bar{W}_1$ has rank n, then $D = K\bar{W}_1^{-1}$. Set $T = \lambda E + D$ where E is a diagonal matrix and $e_{jj} = \delta_{0\bar{j}}$ $(j = 1, 2, \cdots, n)$. The equation $TA = \lambda I - X$ is true for $\lambda = 0$. Therefore $X = -K\bar{W}_1^{-1}A_0$ is a solution of the unilateral equation.

3. The non-commutative case. In this case $\mathfrak{F}$ is taken to be a quasi-field. The polynomial domain $\mathfrak{F}[\lambda]$, where λ is a commutative indeterminate, is a principal ideal ring. Multiplication among the elements is not necessarily commutative, but $d(gf) = d(fg) = d(g) + d(f)$ where g and f are non-zero polynomials.

The above work depends upon the existence of a canonical triangular form for a matrix with elements in $\mathfrak{F}[\lambda]$. V. J. Varineau [4] has shown the existence of a Hermite normal form for a matrix with elements in a non-commutative principal ideal ring. The proof follows the method given by MacDuffee, but is complicated by the lack of determinants. Also, the greatest common right divisor must be used.

---

[5] Published without proof as a note to Ingraham's paper. Ingraham [2].

Once the matrix is reduced to its triangular canonical form, the results of this paper, with the exception of the proof of Theorem 1, may be carried out in the same manner. The concept of rank must be replaced by left row rank or right column rank. The proof of Theorem 1 must be revised when $\mathfrak{F}$ is a quasi-field, as determinants may not be used. The results of Theorem 1 are true, and the reader is referred to the author's doctoral thesis [3] for details of the proof.

Ingraham's algorithm depends upon the existence of a factor theorem for polynomials in λI, with matrix coefficients whose elements are in $\mathfrak{F}$. The commutativity of the elements of the coefficient matrices is not necessary in the proof of the factor theorem. Ore [5] has made a study of congruences of the type (3) over the polynomial domain of a quasi-field. Therefore, the solution of the unilateral matrix equation may be carried out in the same manner as for a commutative field.

In this paper left associates and right unilateral equations have been considered. Similar results may be obtained by dealing with right associates of monic matrices and left unilateral equations. For example, instead of $TA = \lambda^k I + B$ it would be $AT = \lambda^k I + B$. In this event, the canonical triangular form would have the row elements reduced modulo the main diagonal elements. Theory may be developed similar to that already carried out above. In fact, when $\mathfrak{F}$ is a field, the problems are thrown back on the work already done by simply taking the transposes of the matrices involved. For example, $\sum_{m=0}^{r} X^m A_m = 0$ implies $\sum_{m=0}^{r} A^T{}_m (X^T)^m = 0$ (where A^T is the transpose of A).

MICHIGAN STATE COLLEGE,
EAST LANSING, MICHIGAN.

BIBLIOGRAPHY.

[1] C. C. MacDuffee, *The theory of matrices.* Chelsea, 1946.
[2] M. H. Ingraham, "Rational methods in matrix equations," *Bulletin of the American Mathematical Society,* vol. 47 (1941), pp. 61-70.
[3] J. H. Bell, *Topics related to the factorization of matrices.* Submitted as his doctoral dissertation, at the University of Wisconsin. December 1940.
[4] V. J. Varineau, *An extension of the theory of matrices with elements in a principal ideal ring.* This paper was submitted as his doctoral dissertation, at the University of Wisconsin. July 1940.
[5] O. Ore, "Formale Theorie der linearen differential Gleichungen," *Journal für die Reine und Angewandte Mathematik,* vol. 168 (1932), p. 235.

A SUMMABILITY THEOREM FOR DOUBLE ORTHOGONAL SERIES WHOSE COEFFICIENTS SATISFY CERTAIN CONDITIONS.*

By JOSEPHINE MITCHELL.

---

1. Introduction. Let $\{\phi_{mn}(x, y)\}$ $(m, n = 1, 2, \cdots)$ be a complete orthonormal system of real functions of class L^2 defined on the rectangle $Q(a \leq s \leq b,\ c \leq t \leq d)$. By orthonormality we mean that

$$\int_Q \phi_{mn}(s, t)\phi_{pq}(s, t)\,ds\,dt = \delta_{mp}\delta_{nq}, \tag{1.1}$$

where $\delta_{ab} = 0$ if $a \neq b$ and 1 if $a = b$. (All integrals considered in this paper are Lebesgue integrals.) An orthonormal system is complete with respect to functions $f(s, t)$ of class L^2 if

$$\int_Q f(s, t)\phi_{mn}(s, t)\,ds\,dt = 0 \qquad (m, n = 1, 2, \cdots) \tag{1.2}$$

implies $f(s, t) = 0$ almost everywhere on Q.

The orthogonal development of any real function $f(x, y)$ of class L^2 with respect to the system $\{\phi_{mn}(x, y)\}$ is given by

$$\sum_{m,n=1}^{\infty} a_{mn}\phi_{mn}(x, y), \tag{1.3}$$

where

$$a_{mn} = \int_Q f(s, t)\phi_{mn}(s, t)\,ds\,dt \qquad (m, n = 1, 2, \cdots). \tag{1.4}$$

The series (1.3) shall be referred to as the *double orthogonal series* of $f(x, y)$. We shall frequently use the series

$$\sum_{m,n=1}^{\infty} a^2_{mn}. \tag{1.5}$$

We are interested in the connection between the limit of the mn-th partial sum

$$s_{mn} = \sum_{\mu,\nu=1}^{m,n} a_{\mu\nu}\phi_{\mu\nu} \qquad (\phi_{\mu\nu} \equiv \phi_{\mu\nu}(x, y)) \tag{1.6}$$

of the orthogonal series (1.3) and the limit of the Cesàro $(C, 1$

---

* Received October 27, 1947; presented to the American Mat
September 1947.

$$\sigma_{mn} = (1/mn) \sum_{\mu,\nu=1}^{m,n} s_{\mu\nu}. \tag{1.7}$$

For simple orthogonal series it is known that *if series* $\sum_{n=1}^{\infty} a^2{}_n < \infty$, *then* $\sigma_n \rightarrow$ *a limit s almost everywhere in Q if and only if the subsequence* $s_{2^n} \rightarrow s$ *almost everywhere in Q* [3]. In this paper we are able to prove that *if the two series*[1]

$$\text{i. } \sum_{m,n=1}^{\infty} [\log(m+1)]^{1+\epsilon} a^2{}_{mn}, \qquad \text{ii. } \sum_{m,n=1}^{\infty} [\log(n+1)]^{1+\epsilon} a^2{}_{mn}, \tag{1.8}$$

where ϵ *is any arbitrary positive number, both converge, then the double orthogonal series* (1.3) *is* $(C, 1, 1)$ *summable to the sum s almost everywhere in Q if and only if the subsequence* $\{s(2^m, 2^n)\}$ $(s(2^m, 2^n) = s_{2^m,2^n})$ *approaches the limit s almost everywhere in Q as m and* $n \rightarrow \infty$ (Theorem 3.1). For the double series we find it necessary to assume the additional hypotheses concerning series (1.8) in order that certain "cross-product" series may converge (cf. formula (3.4)).

The method of proof used in this paper is a generalization to double orthogonal series of that used for simple orthogonal series. The results on convergence and summability of these series are all stated in [3, Chapter V] but many of the ideas used in our proofs were obtained from a study of numerous papers on convergence and summability of simple orthogonal series to be found in *Fundamenta Mathematicae.* All the theorems proved in this paper hold for higher dimensional spaces but for convenience we state them for the two-dimensional case.

A tool which is used frequently in our proofs is the Schwarz inequality for sums and integrals. For sums this inequality reads

$$(\sum_{i=1}^{N} a_i b_i)^2 \leq \sum_{i=1}^{N} a_i{}^2 \sum_{i=1}^{N} b_i{}^2 \tag{1.9}$$

and similarly for integrals. For the expression "almost everywhere" we use the abbreviation a. e. The symbol $\sum_i \sum_j$ means that we sum in the order $\sum_i (\sum_j)$, while $\sum_{i,j}$ means that any order of summation may be used.

2. An integration theorem. The following theorem, which is a generalization to double sequences of a well-known theorem in Lebesgue integration for simple sequences [3, Chapter I, #2], is necessary for our methods of proof.

---

[1] Logs are taken to the base 2.

THEOREM 2.1. *If the functions* $f_{ij}(s,t)$ $(i,j=1,2,\cdots)\ \varepsilon\ L$ *on* $Q(a\leq s\leq b,\ c\leq t\leq d)$, $f_{ij}(s,t)\leq f_{i'j'}(s,t)$ $(i\leq i',\ j\leq j')$ *a. e. on* Q, *and* $\int_Q f_{ij}(s,t)\,ds\,dt\leq C$ (C *independent of* i *and* j), *then* $\lim_{i,j\to\infty} f_{ij}(s,t)$ *exists a. e. in* Q.

It is often convenient to use Theorem 2.1 as a theorem on series, viz.:

COROLLARY 2.1. *If* $g_{ij}(s,t)\geq 0$ and $\sum_{i,j=1}^{\infty}\int_Q g_{ij}(s,t)\,ds\,dt<\infty$, *then* $\sum_{i,j=1}^{\infty} g_{ij}(s,t)$ *is convergent a. e. in* Q.

3. On the $(C,1,1)$ summability of the double orthogonal series (1.3) if the series (1.8) converge.

3.1. If series (1.8) are assumed to be convergent we can prove the following theorem which corresponds to the result obtained for simple orthogonal series.

THEOREM 3.1. *If series* (1.8) *converge, then a necessary and sufficient condition that the double orthogonal series* (1.3) *be* $(C,1,1)$ *summable to the sum* s *a. e. in* Q *is that the double sequence* $\{s(2^p,2^q)\}$ *converge to* s *a. e. in* Q *as* p *and* $q\to\infty$.

We wish to prove that

$$\lim_{p,q\to\infty}[\sigma(m,n)-s(2^p,2^q)]=0 \text{ a. e. in } Q \tag{3.1}$$

$(2^p<m\leq 2^{p+1},\ 2^q<n\leq 2^{q+1})$ if series (1.8) converge. With this aim in view we break up the difference in (3.1) into

$$\sigma(m,n)-s(2^p,2^q)=[\sigma(m,n)-\sigma(2^p,n)]+[\sigma(2^p,n)-\sigma(2^p,2^q)] + [\sigma(2^p,2^q)-s(2^p,2^q)] \tag{3.2}$$

and prove that each term of the right side of (3.2) converges to 0 under the hypotheses of the Theorem by means of the following set of Lemmas.

3.2. Convergence of a double sum to 0. From definitions (1.6) and (1.7) we have

$$\sigma_{mn}=(1/mn)\sum_{\mu,\nu=1}^{m,n}(m-\mu+1)(n-\nu+1)a_{\mu\nu}\phi_{\mu\nu}. \tag{3.3}$$

Hence

$$(3.4)\qquad \sigma_{mn}-s_{mn}=(1/mn)\sum_{\mu,\nu=1}^{m,n}(\mu-1)(\nu-1)a_{\mu\nu}\phi_{\mu\nu}$$
$$-(1/m)\sum_{\mu,\nu=1}^{m,n}(\mu-1)a_{\mu\nu}\phi_{\mu\nu}-(1/n)\sum_{\mu,\nu=1}^{m,n}(\nu-1)a_{\mu\nu}\phi_{\mu\nu}$$
$$\equiv R_{mn}-P_{mn}-Q_{mn},$$

respectively. For the subsequences $\{2^p\}$ of $\{m\}$ and $\{2^q\}$ of $\{n\}$ it is readily proved by a direct generalization of the one variable methods that $R(2^p,2^q)\to 0$ as p and $q\to\infty$ (See Lemma 3. 1.). The difficulty arises in proving that the sequences $\{P(2^p,2^q)\}$ and $\{Q(2^p,2^q)\}$ converge. However we shall find that if (1. 8) are convergent, then these two sequences $\to 0$ as p and $q\to\infty$.

LEMMA 3. 1. *If series* (1. 5) *is convergent, then*

$$(3.5)\qquad \lim_{p,q\to\infty}R(2^p,2^q)=\lim_{p,q\to\infty}(1/2^p2^q)\sum_{\mu,\nu=1}^{2^p,2^q}(\mu-1)(\nu-1)a_{\mu\nu}\phi_{\mu\nu}=0$$

a. e. in Q [cf. 3, Theorem 585].

Proof. By the orthonormal properties (1. 1) we have formally

$$(3.6)\qquad \sum_{p,q=1}^{\infty}(1/4^p4^q)\int_Q\Big[\sum_{\mu,\nu=1}^{2^p,2^q}(\mu-1)(\nu-1)a_{\mu\nu}\phi_{\mu\nu}\Big]^2ds\,dt$$
$$=\sum_{p,q=1}^{\infty}(1/4^p4^q)\sum_{\mu,\nu=1}^{2^p,2^q}(\mu-1)^2(\nu-1)^2a^2_{\mu\nu}.$$

Since the terms of the latter series are positive we can interchange the order of summation to get[2]

$$(3.7)\qquad \sum_{p,q=1}^{\infty}(1/4^p4^q)\sum_{\mu,\nu=1}^{2^p,2^q}(\mu-1)^2(\nu-1)^2a^2_{\mu\nu}$$
$$=\sum_{\mu,\nu=1}^{\infty}(\mu-1)^2(\nu-1)^2a^2_{\mu\nu}\sum_{p=E[\log\mu]}^{\infty}(1/4^p)\sum_{p=E[\log\nu]}^{\infty}(1/4^q)$$
$$\leq(16/9)\sum_{\mu,\nu=1}^{\infty}a^2_{\mu\nu}.$$

Hence by Corollary 2. 1 the series $\sum_{p,q=1}^{\infty}(1/4^p4^q)\Big[\sum_{\mu,\nu=1}^{2^p,2^q}(\mu-1)(\nu-1)a_{\mu\nu}\phi_{\mu\nu}\Big]^2$ is convergent a. e. in Q, from which equation (3. 5) follows.

3. 3. LEMMA 3. 2. *If series* (1. 8ii) *is convergent, then*

$$(3.8)\qquad \lim_{p,q\to\infty}P(2^p,2^q)=0\ \text{a. e. in}\ Q.$$

---

[2] The notation $E[\log\mu]$ means the nearest integer greater than or equal to $\log\mu$.

Similarly if series (1. 8i) *is convergent, then*

$$\lim_{p,q\to\infty} Q(2^p, 2^q) = 0 \text{ a. e. in } Q. \tag{3.9}$$

Proof. In order to obtain a majorant for the expression $P^2(2^p, 2^q)$ of (3. 4) which shall be monotonic non-decreasing with respect to the subscript 2^q we break up the sum $\sum_{\nu=1}^{2^q}$ in the following manner and apply the Schwarz inequality

$$(\sum_{\nu=1}^{2^q})^2 = (\sum_{\nu=1}^{2} + \sum_{\alpha=2}^{q} \sum_{\nu=2^{\alpha-1}+1}^{2^\alpha})^2 \leq 2(\sum_{\nu=1}^{2})^2 + 2(\sum_{\alpha=2}^{q} \sum_{\nu=2^{\alpha-1}+1}^{2^\alpha})^2. \tag{3.10}$$

Applying this inequality again to the second term on the right we get

$$\begin{aligned}(\sum_{\nu=1}^{2^q})^2 &\leq 2(\sum_{\nu=1}^{2})^2 + 2\sum_{\alpha=2}^{q} (\alpha-1)^{-1-\epsilon} \sum_{\alpha=2}^{q} (\alpha-1)^{1+\epsilon} (\sum_{\nu=2^{\alpha-1}+1}^{2^\alpha})^2 \\ &\leq 2(\sum_{\nu=1}^{2})^2 + 2A_\epsilon \sum_{\alpha=2}^{q} (\alpha-1)^{1+\epsilon} (\sum_{\nu=2^{\alpha-1}+1}^{2^\alpha})^2,\end{aligned} \tag{3.11}$$

where $A_\epsilon = \sum_{n=1}^{\infty} n^{-1-\epsilon} < \infty$ for ϵ any arbitrary positive number. Now let the summand be

$$(1/2^p) \sum_{\mu=1}^{2^p} (\mu-1) a_{\mu\nu} \phi_{\mu\nu} = \Phi_{p\nu} \tag{3.12}$$

and sum with respect to p for $p = 1, 2, \cdots, m$. Then

$$\begin{aligned}\sum_{p=1}^{m} P^2(2^p, 2^q) &\leq 2 \sum_{p=1}^{m} (\sum_{\nu=1}^{2} \Phi_{p\nu})^2 + 2A_\epsilon \sum_{p=1}^{m} \sum_{\alpha=2}^{q} (\alpha-1)^{1+\epsilon} (\sum_{\nu=2^{\alpha-1}+1}^{2^\alpha} \Phi_{p\nu})^2 \\ &= T_{mq},\end{aligned} \tag{3.13}$$

and the sequence $\{T_{mq}\}$ is monotonic non-decreasing in both m and q $(m, q = 1, 2, \cdots)$. Integrate both sides of (3. 13) over the rectangle Q and use the orthonormal properties (1. 1). We get

$$\begin{aligned}\int_Q \sum_{p=1}^{m} P^2(2^p, 2^q)\, ds\, dt &\leq \int_Q T_{mq} ds\, dt \\ &= 2 \sum_{p=1}^{m} \sum_{\nu=1}^{2} (1/4^p) \sum_{\mu=1}^{2^p} (\mu-1)^2 a^2_{\mu\nu} \\ &\quad + 2A_\epsilon \sum_{p=1}^{m} \sum_{\alpha=2}^{q} (\alpha-1)^{1+\epsilon} \sum_{\nu=2^{\alpha-1}+1}^{2^\alpha} (1/4^p) \sum_{\mu=1}^{2^p} (\mu-1)^2 a^2_{\mu\nu}.\end{aligned} \tag{3.14}$$

But $\alpha - 1 = \log 2^{\alpha-1} < \log \nu$ for $\nu = 2^{\alpha-1} + 1, \cdots, 2^\alpha$. Furthermore if we replace $\log \nu$ by $\log(\nu + 1)$, then, since $\log(\nu + 1) \geq 1$ $(\nu = 1, 2, \cdots)$ and

$A_\epsilon > \sum_{n=1}^{\infty} n^{-2} = \pi^2/4 > 1$, we can combine the two inequalities on the right side of (3.14) to obtain

$$(3.15)\qquad \int_Q \sum_{p=1}^{m} P^2(2^p, 2^q)\,ds\,dt \leqq \int_Q T_{mq}\,ds\,dt$$
$$\leqq 2A_\epsilon \sum_{p=1}^{m} \sum_{\nu=1}^{2^q} [\log(\nu+1)]^{1+\epsilon}(1/4^p) \sum_{\mu=1}^{2^p} (\mu-1)^2 a^2_{\mu\nu}.$$

By the same method as we used in inequality (3.7) we can prove that the right side of (3.15) is less than or equal to $(8/3) A_\epsilon \sum_{\mu,\nu=1}^{\infty} [\log(\nu+1)]^{1+\epsilon} a^2_{\mu\nu}$. Hence by Theorem 2.1 the monotonic non-decreasing sequence $\{T_{mq}\}$ approaches a limit a. e. in Q, that is, the series

$$(3.16)\qquad \sum_{p=1}^{\infty} \Big(\sum_{\nu=1}^{2} \Phi_{p\nu}\Big)^2 + \sum_{\substack{p=1\\ \alpha=2}}^{\infty} (\alpha-1)^{1+\epsilon} \Big(\sum_{\nu=2^{\alpha-1}+1}^{2^\alpha} \Phi_{p\nu}\Big)^2$$

is convergent a. e. in Q. Consequently from this result and the absolute convergence of series (3.16), we have

$$(3.17)\qquad \lim_{p,q\to\infty} \Big[\Big(\sum_{\nu=1}^{2} \Phi_{p\nu}\Big)^2 + \sum_{\alpha=2}^{q} (\alpha-1)^{1+\epsilon} \Big(\sum_{\nu=2^{\alpha-1}+1}^{2^\alpha} \Phi_{p\nu}\Big)^2$$
$$= \lim_{p,q\to\infty} \Big[\sum_{\beta=1}^{p} - \sum_{\beta=1}^{p-1}\Big]\Big[\Big(\sum_{\nu=1}^{2} \Phi_{\beta\nu}\Big)^2 + \sum_{\alpha=2}^{q} (\alpha-1)^{1+\epsilon} \Big(\sum_{\nu=2^{\alpha-1}+1}^{2^\alpha} \Phi_{\beta\nu}\Big)^2$$
$$= 0 \text{ a. e. in } Q.$$

But by the inequality expressed in (3.11) with the summand $\Phi_{p\nu}$ of (3.12) the positive term $P^2(2^p, 2^q)$ is less than or equal to $2A_\epsilon$ multiplied by this sum. Hence $\lim_{p,q\to\infty} P^2(2^p, 2^q) = 0$ a. e. in Q and consequently $\lim_{p,q\to\infty} P(2^p, 2^q) = 0$ a. e. in Q. Similarly (3.9) is proved, which completes the proof of Lemma 3.2. From the conclusions of Lemmas 3.1 and 3.2 and equation (3.4) we see that

$$(3.18)\qquad \lim_{p,q\to\infty} [\sigma(2^p, 2^q) - s(2^p, 2^q)] = 0 \text{ a. e. in } Q.$$

Consequently if series (1.8) converge the sequences $\{\sigma(2^p, 2^q)\}$ and $\{s(2^p, 2^q)\}$ are equiconvergent a. e. in Q and to the same limit.

3.4. To complete the proof of Theorem 3.1 we shall prove

LEMMA 3.3. *If series* (1.8i) *converge, then*

$$(3.19)\qquad \lim_{p,q\to\infty} [\sigma(2^p, n) - \sigma(2^p, 2^q)] = 0 \text{ a. e. in } Q,$$

and if series (1.8ii) *converge, then*

$$(3.20)\qquad \lim_{p,q\to\infty} [\sigma(m,n) - \sigma(2^p,n)] = 0 \text{ a. e. in } Q$$

$(2^p < m \leq 2^{p+1}, 2^q < n \leq 2^{q+1})$.

If the integer q be such that $2^q < n \leq 2^{q+1}$, then

$$(3.21)\qquad \sigma(2^p,n) - \sigma(2^p,2^q) = \sum_{\nu=2^q}^{n-1} [\sigma(2^p,\nu+1) - \sigma(2^p,\nu)],$$

and by the Schwarz inequality

$$(3.22)\qquad [\sigma(2^p,n) - \sigma(2^p,2^q)]^2 \leq \sum_{\nu=2^q}^{2^{q+1}-1} (1/\nu) \sum_{\nu=2^q}^{2^{q+1}-1} \nu[\sigma(2^p,\nu+1) - \sigma(2^p,\nu)]^2.$$

Now

$$(3.23)\qquad \sum_{\nu=2^q}^{2^{q+1}-1} (1/\nu) = (1/2^q) + \cdots + (1/2^{q+1}-1) < 2^q/2^q = 1,$$

so that consequently

$$(3.24)\qquad \lim_{p,q\to\infty} [\sigma(2^p,n) - \sigma(2^p,2^q)]^2 \leq \lim_{p,q\to\infty} \sum_{\nu=2^q}^{2^{q+1}-1} \nu[\sigma(2^p,\nu+1) - \sigma(2^p,\nu)]^2$$

and similarly

$$(3.25)\qquad \lim_{p,n\to\infty} [\sigma(m,n) - \sigma(2^p,n)]^2 \leq \lim_{p,n\to\infty} \sum_{\mu=2^p}^{2^{p+1}-1} \mu[\sigma(\mu+1,n) - \sigma(\mu,n)]^2.$$

From inequalities (3.24) and (3.25) Lemma 3.3 will be a consequence of the following Lemma.

LEMMA 3.4. *If series* (1.8i) *is convergent, then*

$$(3.26)\qquad \lim_{m,q\to\infty} \sum_{\nu=2^q}^{2^{q+1}-1} \nu[\sigma(m,\nu+1) - \sigma(m,\nu)]^2 = 0 \text{ a. e. in } Q.$$

Similarly if series (1.8ii) *converges, then*

$$(3.27)\qquad \lim_{p,n\to\infty} \sum_{\mu=2^p}^{2^{p+1}-1} \mu[\sigma(\mu+1,n) - \sigma(\mu,n)]^2 = 0 \text{ a. e. in } Q.$$

Proof. From (3.3) we have that

$$(3.28)\qquad \sigma(m,\nu+1) - \sigma(m,\nu) = \sum_{\alpha,\beta=1}^{m,\nu+1} \frac{m-\alpha+1}{m} \, \frac{\nu-\beta+2}{\nu+1} \, a_{\alpha\beta}\phi_{\alpha\beta} - \sum_{\alpha,\beta=1}^{m,\nu} \frac{m-\alpha+1}{m} \, \frac{\nu-\beta+1}{\nu} \, a_{\alpha\beta}\phi_{\alpha\beta}.$$

But

$$\frac{\nu - \beta + 2}{\nu + 1} - \frac{\nu - \beta + 1}{\nu} = \frac{\beta - 1}{\nu(\nu + 1)} \tag{3.29}$$

so that

$$\sigma(m, \nu + 1) - \sigma(m, \nu) = \sum_{\alpha,\beta=1}^{m,\nu} \frac{(\beta - 1)(m - \alpha + 1)}{\nu(\nu + 1)m} a_{\alpha\beta}\phi_{\alpha\beta} - \sum_{\alpha=1}^{m} \frac{m - \alpha + 1}{m(\nu + 1)} a_{\alpha,\nu+1}\phi_{\alpha,\nu+1} \tag{3.30}$$

and we break up the right side of this expression into the four sums

$$\sum_{\alpha=2}^{m} \sum_{\beta=2}^{\nu} \frac{(\beta - 1)(\alpha - 1)}{\nu(\nu + 1)m} a_{\alpha\beta}\phi_{\alpha\beta} \equiv \sum_{\alpha=2}^{m} \Psi^{(1)}{}_{\alpha\nu}(m) \tag{3.31i}$$

$$\sum_{\alpha=1}^{m} \sum_{\beta=2}^{\nu} \frac{\beta - 1}{\nu(\nu + 1)} a_{\alpha\beta}\phi_{\alpha\beta} \equiv \sum_{\alpha=1}^{m} \Psi^{(2)}{}_{\alpha\nu} \tag{3.31ii}$$

$$\sum_{\alpha=2}^{m} \frac{\alpha - 1}{m(\nu + 1)} a_{\alpha,\nu+1}\phi_{\alpha,\nu+1} \equiv \sum_{\alpha=2}^{m} \psi^{(1)}{}_{\alpha,\nu+1}(m) \tag{3.31iii}$$

$$\sum_{\alpha=1}^{m} \frac{1}{\nu + 1} a_{\alpha,\nu+1}\phi_{\alpha,\nu+1} \equiv \sum_{\alpha=1}^{m} \psi^{(2)}{}_{\alpha,\nu+1}. \tag{3.31iv}$$

We may obtain majorants for the squares of these four sums which shall be monotonic non-decreasing with respect to the subscript m by the following procedure. Let the dyadic representation of m be given by

$$m = 2^p + 1 + \sum_{k=1}^{p} x_k 2^{p-k}, \tag{3.32}$$

where $x_k = 0$ or 1. Using (3.32) we break up the sum $\sum_{\alpha=1}^{m}$ as follows:

$$\sum_{\alpha=1}^{m} = \sum_{\alpha=1}^{2} + \sum_{\mu=2}^{p} \sum_{\alpha=2^{\mu-1}+1}^{2^{\mu}} + \sum_{\mu=0}^{p-1} \sum_{\alpha=s(\mu)}^{s(\mu+1)-1} + \sum_{\alpha=m}, \tag{3.33}$$

where

$$s(\mu) = 2^p + 1 + \sum_{k=1}^{\mu} x_k 2^{p-k} \qquad (\mu = 1, \cdots, p - 1), \tag{3.34}$$
$$s(0) = 2^p + 1.$$

(If $s(\mu) = s(\mu + 1)$ for any μ $(\mu = 0, \cdots, p - 1)$, then the value of the sum $\sum_{\alpha=s(\mu)}^{s(\mu+1)-1}$ will be taken to be zero.)

An illustration of this process for the number $m = 14$ is as follows. The dyadic representation of $14 = 2^3 + 1 + (1.2^2 + 0.2^1 + 1.2^0)$ so that in (3.32), $x_1 = 1$, $x_2 = 0$ and $x_3 = 1$. From (3.34), $s(0) = 9$, $s(1) = 9$

$+ x_1 2^2 = 13$, $s(2) = 9 + x_1 2^2 + x_2 2 = 13$ and $s(3) = 9 + x_1 2^2 + x_2 2 + x_3 2^0 = 14$. Since $s(1) = s(2)$ the sum $\sum_{\alpha=s(1)}^{\alpha=s(2)-1} = 0$. Hence by (3.33) the sum $\sum_{\alpha=1}^{14}$ would be broken up into

$$\sum_{\alpha=1}^{14} = \sum_{\alpha=1}^{2} + \Big(\sum_{\alpha=3}^{4} + \sum_{\alpha=5}^{8}\Big) + \Big(\sum_{\alpha=9}^{12} + \sum_{\alpha=13}\Big) + \sum_{\alpha=14}.$$

By the Schwarz inequality applied to (3.33)

$$(3.35)\qquad \Big(\sum_{\alpha=1}^{m}\Big)^2 \leq 4\Big[\Big(\sum_{\alpha=1}^{2}\Big)^2 + \Big(\sum_{\mu=2}^{p}\sum_{\alpha=2^{\mu-1}+1}^{2^{\mu}}\Big)^2 + \Big(\sum_{\mu=0}^{p-1}\sum_{\alpha=s(\mu)}^{s(\mu+1)-1}\Big)^2 + \Big(\sum_{\alpha=m}\Big)^2\Big],$$

and applying this inequality again to the two inner sums we get

$$(3.36)\qquad \Big(\sum_{\alpha=1}^{m}\Big)^2 \leq 4\Big[\Big(\sum_{\alpha=1}^{2}\Big)^2 + \sum_{\mu=2}^{p}(\mu-1)^{-1-\epsilon}\sum_{\mu=2}^{p}(\mu-1)^{1+\epsilon}\Big(\sum_{\alpha=2^{\mu-1}+1}^{2^{\mu}}\Big)^2 + p\sum_{\mu=0}^{p-1}\Big(\sum_{\alpha=s(\mu)}^{s(\mu+1)-1}\Big)^2 + \Big(\sum_{\alpha=m}\Big)^2\Big]. \quad (\text{cf. } (3.11)).$$

Let the summand in (3.36) be the expression $\Psi^{(1)}{}_{\alpha\nu}(m)$ of (3.31i) and denote the resulting sum on the right side by $t^{(1)}{}_{p\nu}$. Integrating over Q we have

$$(3.37)\qquad \int_Q t^{(1)}{}_{p\nu}\,ds\,dt \leq 4\Big\{A_{2\nu}(m) + \Big[A_\epsilon\sum_{\mu=2}^{p}(\mu-1)^{1+\epsilon}\sum_{\alpha=2^{\mu-1}+1}^{2^{\mu}} + p\sum_{\mu=0}^{p-1}\sum_{\alpha=s(\mu)}^{s(\mu+1)-1}\Big]A_{\alpha\nu}(m) + A_{m\nu}(m)\Big\},$$

where

$$(3.38)\qquad A_{\alpha\nu}(m) = \sum_{\beta=2}^{\mu}\frac{(\beta-1)^2(\alpha-1)^2}{\nu^2(\nu+1)^2 m^2}\,a^2{}_{\alpha\beta}.$$

But $\mu - 1 = \log 2^{\mu-1} < \log\alpha < \log(\alpha+1)$ for $\alpha = 2^{\mu-1}+1, \cdots, 2^{\mu}$ and $p = \log 2^p < \log(\alpha+1)$ for $\alpha = s(\mu)$ to $s(\mu+1)-1$. Hence as in inequality (3.15) we get

$$(3.39)\qquad \begin{aligned}\int_Q t^{(1)}{}_{p\nu}\,ds\,dt &\leq 4\Big\{A_{2\nu}(m) + \Big[A_\epsilon\sum_{\mu=2}^{p}\sum_{\alpha=2^{\mu-1}+1}^{2^{\mu}} + \sum_{\mu=0}^{p-1}\sum_{\alpha=s(\mu)}^{s(\mu+1)-1}\Big]A_{\alpha\nu}(m)[\log(\alpha+1)]^{1+\epsilon} + A_{m\nu}(m)\Big\}\\ &\leq 4A_\epsilon\sum_{\alpha=2}^{m}A_{\alpha\nu}(m)[\log(\alpha+1)]^{1+\epsilon}.\end{aligned}$$

Now $2^p < m \leq 2^{p+1}$. Therefore $(1/m) < (1/2^p) = 2/2^{p+1}$ and using (3.38) we have

$$(3.40)\qquad \int_Q t^{(1)}{}_{p\nu} ds\, dt \leq 16A_\epsilon \sum_{\alpha=2}^{2^{p+1}} \sum_{\beta=2}^{\nu} \frac{(\beta-1)^2(\alpha-1)^2}{\nu^2(\nu+1)^2 4^{p+1}} [\log(\alpha+1)]^{1+\epsilon} a^2{}_{\alpha\beta}.$$

Multiply both sides of this inequality by ν and sum with respect to p and ν. Using the methods of inequality (3.7) and the fact that

$$(3.41)\qquad \sum_{\nu=\beta}^{\infty} (1/\nu^3) \leq 1/\beta^2$$

we get

$$(3.42)\qquad \sum_{p,\nu=1}^{\infty} \nu \int_Q t^{(1)}{}_{p\nu} ds\, dt$$

$$\leq 16A_\epsilon \sum_{p,\nu=1}^{\infty} \sum_{\alpha=2}^{2^{p+1}} \sum_{\beta=2}^{\nu} \frac{(\beta-1)^2(\alpha-1)^2}{\nu^2(\nu+1)^2 4^{p+1}} [\log(\alpha+1)]^{1+\epsilon} a^2{}_{\alpha\beta}$$

$$< 16A_\epsilon \sum_{\alpha,\beta=2}^{\infty} (\beta-1)^2(\alpha-1)^2 [\log(\alpha+1)]^{1+\epsilon} a^2{}_{\alpha\beta} \sum_{p=p'}^{\infty} 4^{-p-1} \sum_{\nu=\beta}^{\infty} \nu^{-3}$$

$$\leq (4/3)(16A_\epsilon) \sum_{\alpha,\beta=2}^{\infty} (\beta-1)^2(\alpha-1)^2 [\log(\alpha+1)]^{1+\epsilon} a^2{}_{\alpha\beta} \alpha^{-2}\beta^{-2}$$

$$\leq (64/3)A_\epsilon \sum_{\alpha,\beta=1}^{\infty} [\log(\alpha+1)]^{1+\epsilon} a^2{}_{\alpha\beta} < \infty,$$

where $p' = E[\log \alpha] - 1$ if $\alpha \neq 2$, $p' = 1$ if $\alpha = 2$. Hence by Corollary 2.1 the series of positive terms, $\sum_{p,\nu=1}^{\infty} \nu t^{(1)}{}_{p\nu}$, is convergent a. e. in Q. Therefore

$$(3.43)\qquad \lim_{p,q\to\infty} \sum_{\nu=2^q}^{2^{q+1}-1} \nu t^{(1)}{}_{p\nu} = \lim_{p,q\to\infty} \Big[\sum_{\nu=1}^{2^{q+1}-1} - \sum_{\nu=1}^{2^q-1}\Big]\Big[\sum_{\alpha=1}^{p} - \sum_{\alpha=1}^{p-1}\Big] \nu t^{(1)}{}_{p\nu} = 0$$

a. e. in Q. Whence by the definition of $t^{(1)}{}_{p\nu}$ and inequality (3.36) it follows that

$$(3.44)\qquad \lim_{m,q\to\infty} \sum_{\nu=2^q}^{2^{q+1}-1} \nu\Big[\sum_{\alpha=2}^{m} \sum_{\beta=2}^{\nu} \frac{(\beta-1)(\alpha-1)}{\nu(\nu+1)m} a_{\alpha\beta}\phi_{\alpha\beta}\Big]^2 = 0 \text{ a. e. in } Q.$$

We may discuss the sum given in (3.31iii) in the same manner as we have done for (3.31i) and it follows that

$$(3.45)\qquad \lim_{m,q\to\infty} \sum_{\nu=2^q}^{2^{q+1}-1} \nu\Big[\sum_{\alpha=2}^{m} \frac{\alpha-1}{m(\nu+1)} a_{\alpha,\nu+1}\phi_{\alpha,\nu+1}\Big]^2 = 0 \text{ a. e. in } Q.$$

In inequality (3.36) used with the summand $\Psi^{(2)}{}_{\alpha\nu}$ of sum (3.31ii) the right side is a monotonic non-decreasing sequence with respect to p. Denote it by $t^{(2)}{}_{p\nu}$. Multiply $t^{(2)}{}_{p\nu}$ by ν and sum with respect to ν for $\nu = 1, \cdots, n$. The resulting sequence $\{T_{pn}\}$ is monotonic with respect to both p and n. Integrating over the rectangle Q we have by the same procedure as used in inequalities (3.37), (3.39) and (3.42) that

$$(3.46)\qquad \int_Q T_{pn}ds\,dt \leq 4A_\epsilon \sum_{\nu=1}^{\infty}\sum_{\alpha=1}^{\infty}\sum_{\beta=2}^{\nu}\frac{(\beta-1)^2}{\nu(\nu+1)^2}[\log(\alpha+1)]^{1+\epsilon}a^2{}_{\alpha\beta}$$

$$\leq 4A_\epsilon \sum_{\alpha=1,\beta=2}^{\infty}(\beta-1)^2[\log(\alpha+1)]^{1+\epsilon}a^2{}_{\alpha\beta}\sum_{\nu=\beta}^{\infty}\nu^{-3}$$

$$\leq 4A_\epsilon \sum_{\alpha,\beta=1}^{\infty}[\log(\alpha+1)]^{1+\epsilon}a^2{}_{\alpha\beta}.$$

Hence by Theorem 2.1 the sequence $\{T_{pn}\}$ approaches a limit a. e. in Q as p and $n \to \infty$. Consequently

$$(3.47)\qquad \lim_{p,q\to\infty}\sum_{\nu=2^q}^{2^{q+1}-1}\nu t^{(2)}{}_p = \lim_{p,q\to\infty}[T_{p,2^{q+1}-1} - T_{p,2^q-1}] = 0 \text{ a. e. in } Q.$$

Therefore from inequality (3.36) used with the summand $\Psi^{(2)}{}_{\alpha\nu}$ it follows that

$$(3.48)\qquad \lim_{m,q\to\infty}\sum_{\nu=2^q}^{2^{q+1}-1}\nu\Big(\sum_{\alpha=1}^{m}\sum_{\beta=2}^{\nu}\frac{\beta-1}{\nu(\nu+1)}a_{\alpha\beta}\phi_{\alpha\beta}\Big)^2 = 0 \text{ a. e. in } Q.$$

The sum (3.31iv) may be discussed in a similar manner and we get that

$$(3.49)\qquad \lim_{m,q\to\infty}\sum_{\nu=2^q}^{2^{q+1}-1}\nu\Big(\sum_{\alpha=1}^{m}\frac{1}{\nu+1}a_{\alpha,\nu+1}\phi_{\alpha,\nu+1}\Big)^2 = 0 \text{ a. e. in } Q.$$

Now from formulas (3.30) and (3.31) and the Schwarz inequality we have that

$$(3.50)\qquad [\sigma(m,\nu+1)-\sigma(m,\nu)]^2$$

$$\leq 4\Big[\Big(\sum_{\alpha=2}^{m}\Psi^{(1)}{}_{\alpha\nu}(m)\Big)^2 + \Big(\sum_{\alpha=1}^{m}\Psi^{(2)}{}_{\alpha\nu}\Big)^2 + \Big(\sum_{\alpha=2}^{m}\psi^{(1)}{}_{\alpha,\nu+1}(m)\Big)^2 + \Big(\sum_{\alpha=1}^{m}\psi^{(2)}{}_{\alpha,\nu+1}\Big)^2\Big].$$

Hence it follows from the conclusions expressed in (3.44), (3.45), (3.48) and (3.49) that if the series (1.8i) converge, then

$$(3.51)\qquad \lim_{m,q\to\infty}\sum_{\nu=2^q}^{2^{q+1}-1}\nu[\sigma(m,\nu+1)-\sigma(m,\nu)]^2 = 0 \text{ a. e. in } Q.$$

Similarly if series (1.8ii) converge, then equation (3.27) may be proved to hold, which completes the proof of Lemma 3.4.

3.5. The following theorem gives a sufficient condition for $(C,1,1)$ summability.

THEOREM 3.2. *If the series* (1.8) *and*

$$\sum_{m,n=1}^{\infty}[\log\log(m+3)]^2[\log\log(n+3)]^2a^2{}_{mn}$$

all converge, then the double orthogonal series (1.3) *is* $(C, 1, 1)$ *summable a. e. in the rectangle* $Q(a \leq x \leq b, c \leq y \leq d)$.

Proof. This theorem is a consequence of Theorem 3.1 and the following theorem which is due to R. P. Agnew [1, Theorem 16.1].

THEOREM. *If the coefficients in a double orthogonal series are such that*

$$\sum_{m,n=1}^{\infty} [\log\log(m+3)]^2[\log\log(n+3)]^2 a^2_{mn}$$

is convergent, then there exists a function $f(x, y)$ *such that* $\lim_{p,q\to\infty} s(2^p, 2^q) = f(x, y)$ *essentially uniformly over* Q.[3]

Theorem 4.3 [1] would also give a sufficient condition for $(C, 1, 1)$ summability of the double orthogonal series.

UNIVERSITY OF ILLINOIS.

REFERENCES.

[1] R. P. Agnew, "On double orthogonal series," *Proceedings of the London Mathematical Society*, vol. 33 (1932), pp. 420-434.

[2] T. J. Ia Bromwich, *An Introduction to the Theory of Infinite Series*, Macmillan and Co., 2nd edition (1931).

[3] S. Kaczmarz and H. Steinhaus, *Theorie der Orthogonalreihen*, Monografje Matematyczne, Tom VI, Warsaw (1935).

---

[3] Essentially uniform convergence over Q implies convergence a. e. over Q.

NOTE ON THE PROPERTIES OF FOURIER COEFFICIENTS.*

By Ching-Tsün Loo.

---

1. Introduction. In this paper we shall consider the Fourier series of real-valued and L-integrable functions of period 2π. We shall often consider only even or odd functions. Even functions will be denoted by $f(x)$, and their cosine Fourier coefficients by a_n; odd functions will be denoted by $g(x)$ and their sine coefficients by b_n. Thus

$$(1.1) \qquad f(x) \sim \sum_{n=1}^{\infty} a_n \cos nx, \qquad g(x) \sim \sum_{n=1}^{\infty} b_n \sin nx,$$

assuming for simplicity that the integral of $f(x)$ over $(0, 2\pi)$ is zero. We shall write

$$(1.2) \qquad A_n = 1/n \sum_{k=1}^{n} a_k, \qquad A^*_n = \sum_{k=n}^{\infty} a_k/k,$$

and similarly we define B_n, B^*_n. If the series

$$(1.3) \qquad \sum_{n=1}^{\infty} A_n \cos nx, \qquad \sum_{n=1}^{\infty} A_n \sin nx, \qquad \sum_{n=1}^{\infty} A^*_n \cos nx, \qquad \sum_{n=1}^{\infty} A^*_n \sin nx$$

are Fourier series, the functions they represent will be denoted by F_c, F_s, F^*_c, F^*_s respectively. Similarly we have G_c, G_s, G^*_c, G^*_s.

Hardy [2] proved that if $f(x) \varepsilon L^p$ $(1 \leq p < \infty)$, then $\sum_{n=1}^{\infty} A_n \cos nx$ is a Fourier series of the class L^p. Recently, Bellman [1] proved[1] that if $f(x) \varepsilon L^p$ $(1 < p < \infty)$ then $\sum_{n=1}^{\infty} A^*_n \cos nx$ is a Fourier series of class L^p. Here the restriction to $p > 1$ is essential, since for $p = 1$ the numbers A^*_n need not exist, as the familiar example of $f(x) \sim \sum_{n=2}^{\infty} \cos nx/\log n$ shows.

On account of the well-known theorems of M. Riesz (according to which the series conjugate to a series of the class L^p $(1 < p < \infty)$ is also of the class L^p) the results of Hardy and Bellman show that if $f(x) \varepsilon L^p$

---

* Received January 27, 1948.

[1] The result of Bellman was also obtained by Kawata and Sunouchi [3], [4]. These two papers came to my knowledge only very recently after the results of the present paper had been obtained. Though the Kawata-Sunouchi theorems have points in common with my results, none of the theorems established here is proved in their papers.

$(1 < p < \infty)$ then all the series in (1.3) are Fourier series of the class L^p. The arguments of Hardy and Bellman can also be applied to show that if $g(x) \varepsilon L^p$ $(1 \leq p < \infty)$ then $\sum_{n=1}^{\infty} B_n \sin nx$ is a Fourier series, and if $g(x) \varepsilon L^p$ $(1 < p < \infty)$ then $\sum_{n=1}^{\infty} B^*_n \sin nx$ is a Fourier series, both of the class L^p. By making use of M. Riesz's theorem we see that if $g(x) \varepsilon L^p$ $(1 < p < \infty)$ then the four series obtained by replacing the a_n in (1.3) by b_n are also Fourier series of the class L^p. The purpose of this paper is to investigate the cases $p = 1$ and $p = \infty$. They are delicate and interesting.

2. First, we shall consider the four cases corresponding to the result of Hardy. We formulate Hardy's theorem in the case of $p = 1$.

THEOREM 1. *If $f(x) \varepsilon L$ then $\sum_{n=1}^{\infty} A_n \cos nx$ is a Fourier series.*

By making use of Hardy's analysis one can also prove

THEOREM 2. *If $g(x) \varepsilon L$ then $\sum_{n=1}^{\infty} B_n \sin nx$ is a Fourier series.*

It is interesting to observe that in general, the condition $f(x) \varepsilon L$ does not imply the Fourier character of $\sum_{n=1}^{\infty} A_n \sin nx$. For $\sum_{n=2}^{\infty} \cos nx/\log n$ is the Fourier series of an even and integrable function but the corresponding $A_n = 1/n \sum_{\nu=2}^{n} 1/\log \nu \simeq 1/\log n$ and $\sum_{n=2}^{\infty} \sin nx/\log n$ is not a Fourier series.[2] In this case we see that $\sum_{n=1}^{\infty} a_n \sin nx$ is a series conjugate to $f(x) \sim \sum_{n=1}^{\infty} a_n \cos nx$. It certainly represents a Fourier series of $\bar{f}(x)$ if $| f | \log^+ | f | \varepsilon L$ (a theorem of Zygmund: if $| h | \log^+ | h | \varepsilon L$ then $\bar{h} \varepsilon L$, where h is not necessarily confined to be even or odd, and $\bar{h}$ is the conjugate function of h). We also observe that $\sum_{n=1}^{\infty} a_n \sin nx$ is a $g(x)$ with $a_n = b_n$. By using Theorem 2, we conclude that $\sum_{n=1}^{\infty} A_n \sin nx$ is a Fourier series. Thus we have

THEOREM 3. *If $| f | \log^+ | f | \varepsilon L$ then $\sum_{n=1}^{\infty} A_n \sin nx$ is a Fourier series.*

Using Theorem 1 and applying a similar argument we can deduce

---

[2] Of course it does not follow from the fact that $\beta_n \simeq \beta'_n$ that the behavior of the series $\Sigma \beta_n \sin nx$ and $\Sigma \beta'_n \sin nx$ is exactly the same. However in our example (and in similar examples that follow) the assertion is justified by the very regular structure of the sequences $\{\beta_n \mid \beta'_n\}$, $\{\beta'_n \mid \beta_n\}$. We omit the proofs which can easily be supplied by the reader (see for example Zygmund [6]).

THEOREM 4. *If* $|g| \log^+ |g| \varepsilon L$ *then* $\sum_{n=1}^{\infty} B_n \cos nx$ *is a Fourier series.*

We shall now consider the four cases corresponding to the result of Bellman. To make the situation clear, we note the following examples. We see that $f(x) \sim \sum_{n=2}^{\infty} \cos nx/\log n$ is even and integrable but $A^*_n = \sum_{\nu=n}^{\infty} 1/\nu \log \nu$ does not exist. This means that the series $\sum_{n=2}^{\infty} A^*_n \cos nx$ and $\sum_{n=2}^{\infty} A^*_n \sin nx$ in general have no meaning. We also see that $g(x) \sim \sum_{n=2}^{\infty} \sin nx/(\log n)^2$ is odd and integrable but $B^*_n = \sum_{\nu=2}^{n} 1/\nu(\log \nu)^2 \simeq 1/\log n$ and $\sum_{n=2}^{\infty} \sin nx/\log n$ is not a Fourier series. However, the following theorems can be established:

THEOREM 5. *If* $|g| \log^+ 1/|x| \varepsilon L$ *then* $\sum_{n=1}^{\infty} B^* \sin nx$ *is a Fourier series.*

THEOREM 6. *If* $|f| \log^+ 1/|x| \varepsilon L$ *then* $\sum_{n=1}^{\infty} A^*_n \cos nx$ *is a Fourier series.*[3]

THEOREM 7. *If* $|g| \log^+ |g| \varepsilon L$ *then* $\sum_{n=1}^{\infty} B^*_n \cos nx$ *is a Fourier series.*

It is interesting to observe that the condition $|f| \log^+ |f| \varepsilon L$ does not imply the Fourier character of $\sum_{n=1}^{\infty} A^*_n \sin nx$ as shown by the example $f(x) \sim \sum_{n=2}^{\infty} \cos nx/(\log n)^2$. From Theorem 5 and the theorem of Zygmund that if $|h| (\log^+ |h|)^\alpha \varepsilon L$, $\alpha > 0$, then $|\bar{h}| (\log |\bar{h}| + 2)^{\alpha-1} \varepsilon L$, we can deduce the following

THEOREM 8. *If* $|f| \log^2 |f| \varepsilon L$ *then* $\sum_{n=1}^{\infty} A^*_n \sin nx$ *is a Fourier series.*

Indeed, under the condition $|f| \log^2 |f| \varepsilon L$ we have

$$f(x) \sim \sum_{n=1}^{\infty} a_n \cos nx, \qquad \bar{f}(x) \sim \sum_{n=1}^{\infty} a_n \sin nx$$

and $|\bar{f}| \log (|\bar{f}| + 2) \varepsilon L$. Since $\bar{f}(x)$ is a $g(x)$ with $a_n = b_n$, it follows from Theorem 5 that $\sum_{n=1}^{\infty} A^*_n \sin nx$ is a Fourier series.

3. This section is devoted to the proofs of Theorems 5, 6, 7. Following Hardy, we modify the definitions of A_n, A^*_n, B_n, B^*_n, and write

---

[3] The condition given in Theorems 5 and 6 is less stringent than $|h| \log^+ |h| \varepsilon L$ (Zygmund [6], p. 106, ex. 4, 5.)

$$A_n = 1/n \sum_{\nu=1}^{n} a_\nu - a_n/2n, \qquad A^*_n = \sum_{\nu=n+1}^{\infty} a_\nu/\nu + a_n/2n,$$

with similar formulas for B_n and B^*_n. With these new definitions one gets similar formulas, and the difference between the old and new A_n (or A^*_n, etc.) being $o(1/n)$, the difference of the corresponding functions is a function integrable in any power (by a theorem of Hausdorff-Young) or even exponentially integrable (see Zygmund [5]). We shall use two well-known functions. The odd function $\frac{1}{2}(\pi - x) \sim \sum_{n=1}^{\infty} \sin nx/n$ $(0 < x \leq \pi)$ and the even function $\log 1/2 \sin \frac{1}{2}x \sim \sum_{n=1}^{\infty} \cos nx/n$ $(0 < x \leq \pi)$. The method used is to find explicit formulas for the functions with A^*_n and B^*_n as Fourier coefficients.

Let

$$\alpha(x) = \tfrac{1}{2}(\pi \operatorname{sgn} x - x) \sim \sum_{n=1}^{\infty} \sin nx/n. \tag{3.1}$$

It is easy to see that

$$\sum_{\nu=1}^{\infty} b_\nu/\nu = 1/\pi \int_{-\pi}^{\pi} g(t)\alpha(t)\,dt, \tag{3.2}$$

and that

$$\sum_{\nu=1}^{\infty} b_\nu \cos \nu x/\nu = 1/\pi \int_{-\pi}^{\pi} g(t)\alpha(x+t)\,dt.$$

Hence

$$b_\nu/\nu = 2/\pi^2 \int_0^{\pi} \cos \nu x \int_{-\pi}^{\pi} g(t)\alpha(x+t)\,dt\,dx,$$

and (taking into account that the integral of $\sum_{\nu=1}^{\infty} b_\nu \nu^{-1} \cos \nu x$ over $(0, \pi)$ is zero)

$$\sum_{\nu=1}^{n} b_\nu/\nu - b_n/2n = 1/\pi^2 \int_0^{\pi} \sin nx \cot \tfrac{1}{2}x \int_{-\pi}^{\pi} g(t)\alpha(x+t)\,dt\,dx.$$

Subtracting the last formula from (3.2), we thus obtain

$$\begin{aligned} B^*_n &= \sum_{\nu=n+1}^{\infty} b_\nu/\nu + b_n/2n \\ &= 1/\pi^2 \int_0^{\pi} \sin nx \cot \tfrac{1}{2}x \int_{-\pi}^{\pi} g(t)[\alpha(t) - \alpha(x+t)]\,dt\,dx \\ &= 1/\pi^2 \int_0^{\pi} \sin nx \cot \tfrac{1}{2}x \int_0^{\pi} g(t)[2\alpha(t) - \alpha(x+t) + \alpha(x-t)]\,dt\,dx. \end{aligned}$$

It remains to establish the integrability of the function

$$\frac{1}{2\pi} \cot \tfrac{1}{2}x \int_0^{\pi} g(t)M(x,t)\,dt \tag{3.3}$$

where

$$M(x,t) = 2\alpha(t) - \alpha(x+t) + \alpha(x-t). \tag{3.4}$$

It is easy to see that for x and t both in the interval $(0,\pi)$,

$$M(x,t) = \tfrac{1}{2}\pi(1+\operatorname{sgn}(x-t)) = \begin{cases} 0 & (t > x), \\ \pi & (t < x). \end{cases} \tag{3.5}$$

This follows immediately from the definition of $\alpha(x)$. Hence

$$\begin{aligned} I &= \int_0^\pi \cot \tfrac{1}{2}x \int_0^\pi g(t) M(x,t)\, dtdx \\ &= \pi \int_0^\pi \cot \tfrac{1}{2}x \int_0^x g(t)\, dtdx, \end{aligned}$$

and

$$\begin{aligned} |I| \leq 2\pi \int_0^\pi 1/x \int_0^x |g(t)|\, dtdx &= 2\pi \int_0^\pi |g(t)| \int_t^\pi (1/x)\, dxdt \\ &= 2\pi \int_0^\pi |g(t)|(\log \pi/t)\, dt. \end{aligned}$$

Theorem 5 is thus proved.

In order to prove Theorem 6, we argue as before and we get

$$\begin{aligned} \sum_{\nu=1}^{n} a_\nu/\nu - a_n/2n &= 1/\pi^2 \int_0^\pi (1-\cos nx) \cot \tfrac{1}{2}x \int_{-\pi}^\pi f(t)\alpha(x+t)\, dtdx \\ &= 1/\pi^2 \int_0^\pi (1-\cos nx) \cot \tfrac{1}{2}x \int_0^\pi f(t)[\alpha(x+t) + \alpha(x-t)]\, dtdx. \end{aligned} \tag{3.6}$$

It suffices to establish the integrability of

$$\tfrac{1}{2\pi} \cot \tfrac{1}{2}x \int_0^\pi f(t) N(x,t)\, dt, \tag{3.7}$$

where

$$N(x,t) = \alpha(x+t) + \alpha(x-t), \tag{3.8}$$

since then

$$\int_0^\pi \cos nx \cot \tfrac{1}{2}x \int_0^\pi f(t) N(x,t)\, dtdx \to 0$$

as $n \to \infty$ by the Riemann-Lebesgue theorem. Thus

$$\sum_{\nu=1}^{\infty} a_\nu/\nu = 1/\pi^2 \int_0^\pi \cot \tfrac{1}{2}x \int_0^\pi f(t) N(x,t)\, dtdx,$$

by (3. 6), and therefore $A^*_n = \sum_{\nu=n+1}^{\infty} a_\nu/\nu + a_n/2n$ is the Fourier cosine coefficient of (3. 7).

We observe that for x and t both in the interval $(0, \pi)$,

$$(3.9) \qquad N(x, t) = \tfrac{1}{2}\pi(1 + \operatorname{sgn}(x - t)) - x = \begin{cases} -x & (t > x), \\ \pi - x & (t < x). \end{cases}$$

Hence

$$\int_0^\pi \cot \tfrac{1}{2}x \int_0^\pi f(t) N(x, t)\, dt dx$$
$$= \int_0^\pi \cot \tfrac{1}{2}x \int_0^x (\pi - x) f(t)\, dt dx - \int_0^\pi x \cot \tfrac{1}{2}x \int_x^\pi f(t)\, dt dx.$$

The absolute value of the second integral is majorized by $2\pi \int_0^\pi | f(t) |\, dt$, while the absolute value of the first integral is majorized by

$$2\pi \int_0^\pi 1/x \int_0^x | f(t) |\, dt dx = 2\pi \int_0^\pi | f(t) | \int_t^\pi (1/x)\, dx dt$$
$$= 2\pi \int_0^\pi | f(t) | \log (\pi/t)\, dt.$$

Theorem 6 is thus proved.

The method used to prove Theorem 7 is similar. We shall need the following

LEMMA. *Let* $\theta(x, | h |) = \theta(x, | h |, a, b) = \sup_\xi 1/(x - \xi) \int_\xi^x | h(t) |\, dt.$ *If* $| h | \log^+ | h | \,\varepsilon\, L(a, b)$, *then* $\theta(x, | h |) \,\varepsilon\, L(a, b)$ *and*

$$\int_a^b \theta(x, | h |)\, dx \leq B \int_a^b | h | \log^+ | h |\, dx + C$$

where B *and* C *depend on* $b - a$ *only.* (cf. Zygmund [6]).

Since

$$g(x) \sim \sum_{n=1}^\infty b_n \sin nx, \qquad \log 1/2 \,|\sin \tfrac{1}{2}x| \sim \sum_{n=1}^\infty \cos nx/n$$

we have

$$-\sum_{n=1}^\infty b_n \sin nx/n = 1/\pi \int_{-\pi}^\pi g(t)\ (\log 1/2 \,| \sin (x + t)/2 |)\, dt.$$

It follows that

$$b_\nu/\nu = -2/\pi^2 \int_0^\pi \sin \nu x \int_{-\pi}^\pi g(t)\ (\log 1/2 \,| \sin (x + t)/2 |)\, dt dx$$

and that

$$\sum_{\nu=1}^n b_\nu/\nu - b_n/2n$$
$$= -1/\pi^2 \int_0^\pi (1 - \cos nx) \cot \tfrac{1}{2}x \int_{-\pi}^\pi g(t) (\log 1/2 \,| \sin (x + t)/2 |)\, dt dx$$
$$= -1/\pi^2 \int_0^\pi (1 - \cos nx) \cot \tfrac{1}{2}x \int_0^\pi g(t) \log |(\sin(x - t)/2)/(\sin(x + t)/2)|\, dt$$

As in the proof of Theorem 6, it suffices to prove that

$$(3.10)\qquad -1/2\pi\cot\tfrac{1}{2}x\int_0^{\pi} g(t)\log|(\sin(x-t)/2)/(\sin(x+t)/2)|\,dt$$

is integrable. Let us write $G(x)=\int_0^x g(t)\,dt$. Since $|g|\log^+|g|$ is integrable, the integral

$$(3.11)\qquad \int_0^{\pi} g(t)\log|(\sin(x-t)/2)/(\sin(x+t)/2)|\,dt$$

exists (see Zygmund, [6], p. 106, ex. 4). Moreover, it is not difficult to see that $G(x+\epsilon)-G(x)=o(1/|\log\epsilon|)$ as $\epsilon\to+0$ (see Zygmund [6], p. 105, ex. 3; the general result stated there only gives $G(x+\epsilon)-G(x)=O(1/|\log\epsilon|)$ but the passage from 'O' to 'o' is simple). Thus if we split the integral (3.11) into the integrals over (o,x) and (x,π), integrate each by parts and add the results we find that (3.11) equals

$$\{(G(t)-G(x))\log|(\sin(x-t)/2)/(\sin(x+t)/2)|\Big|_0^{\pi}\}$$
$$+\int_0^{\pi}\frac{(G(t)-G(x))\sin x}{2\sin(x-t)/2\sin(x+t)/2}\,dt$$
$$=\int_0^{\pi}\frac{(G(t)-G(x))\sin x}{2\sin(x-t)/2\sin(x+t)/2}\,dt.$$

In order to prove the integrability of (3.10), it suffices to prove that

$$(3.12)\qquad \int_0^{\pi}\tfrac{1}{2}\cot\tfrac{1}{2}x\{\int_0^{\pi-x}+\int_{\pi-x}^{\pi}\}\frac{|G(t)-G(x)|}{\sin(|x-t|/2)\sin(x+t)/2}\,dtdx$$
$$=J_1+J_2$$

is bounded.

It is easy to see that

$$J_1\leq\int_0^{\pi}\int_0^{\pi-x}\frac{|G(t)-G(x)|}{\sin(|x-t|/2)\sin(x+t)/2}\,dtdx.$$

If we notice that $\frac{1}{2}|x-t|$ and $\frac{1}{2}(x+t)$ are both positive and not greater than $\pi/2$, we can use the formula $\sin\theta\geq(2/\pi)\theta$ in $(0\leq\theta\leq\pi/2)$. Thus

$$(3.13)\qquad J_1\leq\pi^2\int_0^{\pi}\int_0^{\pi}\frac{|G(t)-G(x)|}{|x-t|(x+t)}\,dtdx$$
$$=\pi^2\int_0^{\pi}\{\int_0^{x}+\int_x^{\pi}\}\frac{|G(t)-G(x)|}{|x-t|(x+t)}\,dtdx=J'_1+J''_1$$

say. Now

$$J'_1 \leq \pi^2 \int_0^\pi 1/x \int_0^x (|G(x) - G(t)|/(x-t))\,dtdx$$
$$\leq \pi^2 \int_0^\pi \theta(x, |g|)/x \int_0^x dtdx = \pi^2 \int_0^\pi \theta(x, |g|)\,dx,$$

and

$$J''_1 \leq \pi^2 \int_0^\pi \int_x^\pi (|G(t) - G(x)|/(t-x)t)\,dtdx$$
$$\leq \pi^2 \int_0^\pi \int_x^\pi (\theta(t, |g|)/t)\,dtdx$$
$$= \pi^2 \int_0^\pi \theta(t, |g|)/t \int_0^t dxdt = \pi^2 \int_0^\pi \theta(t, |g|)\,dt,$$

and it follows that

$$J_1 \leq J'_1 + J''_1 \leq 2\pi^2 \int_0^\pi \theta(x, |g|)\,dx. \tag{3.14}$$

Also,

$$J_2 \leq \int_0^{\pi/2} \int_{\pi-x}^\pi \frac{|G(t) - G(x)|}{\sin(|x-t|/2)\sin(x-t)/2}\,dtdx \tag{3.15}$$
$$+ \int_{\pi/2}^\pi \tfrac{1}{2}\cot\tfrac{1}{2}x \int_{\pi-x}^\pi \frac{|G(t) - G(x)|\sin x}{\sin(|x-t|/2)\sin(x-t)/2}\,dtdx$$
$$= J'_2 + J''_2.$$

To treat J'_2, we notice that $\pi/2 \leq (x+t)/ \leq 3\pi/4$, so that $\sin(x+t)/2 \geq \sin 3\pi/4 = 1/\sqrt{2}$, and

$$J'_2 \leq \sqrt{2}\,\pi \int_0^\pi \int_0^\pi (|G(t) - G(x)|/|x-t|)\,dtdx \leq 2\sqrt{2}\,\pi^2 \int_0^\pi \theta(x, |g|)\,dx.$$

Lastly, in J''_2 we have $|\cot\frac{1}{2}x| \leq 1$ and $\sin(x+t)/2 \geq \sin(x+\pi)/2 = \cos x/2$, since $\pi/2 \leq (x+t)/2 \leq \pi$. Therefore

$$J''_2 \leq \pi \int_0^\pi \int_0^\pi \frac{|G(t) - G(x)|\sin x}{2|x-t|\cos x/2}\,dtdx$$
$$\leq \pi \int_0^\pi \int_0^\pi \frac{|G(t) - G(x)|}{|x-t|}\,dtdx \leq 2\pi^2 \int_0^\pi \theta(x, |g|)\,dx.$$

Hence

$$J_2 \leq J'_2 + J'' \leq 6\pi^2 \int_0^\pi \theta(x, |g|)\,dx. \tag{3.16}$$

Combining (3.12), (3.14), (3.16) we find that the integral in (3.12) is certainly bounded by $8\pi^2 \int_0^\pi \theta(x, |g|)\,dx$. An application of the Lemma stated above completes the proof of Theorem 7.

4. In this section we shall always assume that our functions $f(x)$ and $g(x)$ do not exceed 1 in absolute value. Under this condition we shall investigate the integrability of $\exp \lambda \mid F_c \mid$, $\exp \lambda \mid G_c \mid$, etc.

We first recall the proof of Theorems 5 and 6 in **3.** We see that if $g(x) \sim \sum_{n=1}^{\infty} b_n \sin nx$ then

$$(4.1) \qquad G^*_s = \frac{1}{2\pi} \cot \tfrac{1}{2}x \int_0^{\pi} g(t) M(x, t)\, dt \sim \sum_{n=1}^{\infty} B^*_n \sin nx,$$

where $M(x, t)$ is defined by (3.4) and has the properties (3.5). We also see that if $f(x) \sim \sum_{n=1}^{\infty} a_n \cos nx$, then

$$(4.2) \qquad F^*_c = \frac{1}{2\pi} \cot \tfrac{1}{2}x \int_0^{\pi} f(t) N(x, t)\, dt \sim \sum_{n=1}^{\infty} A^*_n \cos nx,$$

where $N(x, t)$ is defined by (3.8) and has the properties (3.9). We easily deduce from (4.1) and (4.2) the following results.

THEOREM 9. *If* $| g(x) | \leq 1$, *then* $| G^*_s(x) | \leq 1$.

THEOREM 10. *If* $| f(x) | \leq 1$, *then* $| F^*_c(x) | \leq 2$.

From the above two theorems, from the theorem that if $| h(x) | \leq 1$, then $\exp \lambda \mid h(x) \mid$ is integrable for every $\lambda < \pi/2$ (Cf. [5]), and from the facts that $G^*_c(x)$ is a $\bar{G}^*(x)$, and $F^*_s(x)$ is an $\bar{F}^*_c(x)$, we thus obtain the next two theorems.

THEOREM 11. *If* $| g(x) | \leq 1$, *then*

$$\int_0^{\pi} \exp \lambda \mid G^*_c(x) \mid dx < \infty$$

for every $\lambda < \frac{1}{2}\pi$.

THEOREM 12. *If* $| f(x) | \leq 1$, *then*

$$\int_0^{\pi} \exp \lambda \mid F^*_s(x) \mid dx < \infty$$

for every $\lambda < \frac{1}{4}\pi$.

In order to deduce the corresponding properties of $\exp \lambda \mid F_c(x) \mid$, $\exp \lambda \mid F_s(x) \mid$, $\exp \lambda \mid G^*_c(x) \mid$, $\exp \lambda \mid G^*_s(x) \mid$, we shall make use of an argument from Hardy's paper [2]. He proved there that if $f(x) \sim \sum_{n=1}^{\infty} a_n \cos nx$ then

$$(4.3) \qquad F_c(x) = \tfrac{1}{2} \int_x^{\pi} \cot \tfrac{1}{2}u f(u)\, du \sim \sum_{n=1}^{\infty} A_n \cos nx.$$

In the same way, we can prove that if $g(x) \sim \sum_{n=1}^{\infty} b_n \sin nx$ then

$$(4.4) \qquad G_c(x) = \tfrac{1}{2} \int_x^{\pi} \cot \tfrac{1}{2} u g(u)\, du \sim \sum_{n=1}^{\infty} B_n \sin nx.$$

From (4.3) and (4.4) and from the inequality

$$(4.5) \qquad \left| \tfrac{1}{2} \int_x^{\pi} \cot \tfrac{1}{2} u f(u)\, du \right| \leq \int_x^{\pi} du/u = \log \pi/x,$$

we obtain

Theorem 13. *If* $|f(x)| \leq 1$, *then*

$$\int_0^{\pi} \exp \lambda \, |F_c(x)| \, dx < \infty$$

for every $\lambda < 1$.

Theorem 14. *If* $|g(x)| \leq 1$, *then*

$$\int_0^{\pi} \exp \lambda \, |G_s(x)| \, dx < \infty$$

for every $\lambda < 1$.

In order to deduce further results we require the following

Theorem 15. *If* $|h| \leq 1$, *then*

$$(4.6) \qquad \Lambda = \tfrac{1}{2} \left| \int_x^{\pi} \bar{h}(t) \cot \tfrac{1}{2} t \, dt \right| \leq 1/\pi \log^2 1/x + O(\log 1/x),$$

where $\bar{h}(t) = 1/\pi \int_{-\pi}^{\pi} (h(u)/2 \tan \tfrac{1}{2}(t-u))\, du$ *is the conjugate function of* $h(t)$.

Since its proof is rather long, we shall postpone it to **5**.

If we start with a bounded odd function $g(x) \sim \sum b_n \sin nx$, then $\bar{g}(x) \sim \sum b_n \cos nx$. Observing that $\bar{g}(x)$ is an $f(x)$ with $b_n = a_n$, and making use of (4.3), we have

$$(4.7) \qquad G_c(x) = \tfrac{1}{2} \int_x^{\pi} \bar{g}(t) \cot \tfrac{1}{2} t \, dt \sim \sum_{n=1}^{\infty} B_n \cos nx.$$

Using Theorem 15 we thus obtain

Theorem 16. *If* $|g(x)| \leq 1$, *then*

$$\int_0^{\pi} \exp \lambda \, |G_c(x)|^{1/2} dx < \infty$$

for every $\lambda < \pi^{1/2}$.

With a similar argument, we obtain

THEOREM 17. *If* $|f(x)| \leq 1$, *then*

$$\int_0^\pi \exp \lambda \, |F_s(x)|^{1/2} dx < \infty$$

for every $\lambda < \pi^{1/2}$.

5. Proof of Theorem 15. We shall estimate the integral

$$\Lambda = \tfrac{1}{2\pi} \left| \int_x^\pi \cot \tfrac{1}{2} t \int_{-\pi}^{\pi} (h(u)/2 \tan \tfrac{1}{2}(t-u))\, du dt \right| . \tag{5.1}$$

We notice that we can interchange the order of integration here (for the inner integral is a Cauchy integral). If we let

$$\begin{aligned} \bar{f}_\epsilon(t) &= 1/\pi \Big(\int_{-\pi}^{t-\epsilon} + \int_{t+\epsilon}^{\pi+t} \Big) (h(u)/2 \tan \tfrac{1}{2}(t-u))\, du \\ &= -1/\pi \int_\epsilon^\pi (h(t+u) - h(t-u)/2 \tan \tfrac{1}{2} u)\, du, \end{aligned}$$

then by a theorem of Zygmund (cf. [6] pp. 249-250), the function $|\bar{f}_\epsilon(t)|$ is certainly bounded by an integrable function independent of ϵ. Hence by the very well-known theorem of Lebesgue about the termwise integration of sequences of functions majorized by integrable functions,

$$\begin{aligned} &\tfrac{1}{2\pi} \int_x^\pi \cot \tfrac{1}{2} t \lim_{\epsilon \to +0} \Big(\int_{-\pi}^{t-\epsilon} + \int_{t+\epsilon}^{\pi} \Big) (h(u)/2 \tan \tfrac{1}{2}(t-u))\, du dt \\ &\qquad = \lim_{\epsilon \to +0} \tfrac{1}{2\pi} \int_x^\pi \cot \tfrac{1}{2} t \Big(\int_{-\pi}^{t-\epsilon} + \int_{t+\epsilon}^{\pi} \Big) (h(u)/2 \tan \tfrac{1}{2}(t-u))\, du dt. \end{aligned}$$

We now interchange the order of integration and let $\epsilon \to +0$, then we get

$$\tfrac{1}{2\pi} \int_{-\pi}^{\pi} h(u) \int_x^\pi \tfrac{1}{2} \cot \tfrac{1}{2} t \cot \tfrac{1}{2}(t-u)\, dt du. \tag{5.2}$$

Using the identity

$$\cot t \cot (t-u) = \cot u (\cot (t-u) - \cot t) - 1,$$

we reduce the above integral (5.2) to

$$\tfrac{1}{2\pi} \int_{-\pi}^{\pi} h(u) \cot u/2 \int_x^\pi \tfrac{1}{2} [\cot \tfrac{1}{2}(t-u) - \cot \tfrac{1}{2} t]\, dt du + O(1). \tag{5.3}$$

In order to establish (4.6), we have only to prove that

$$(5.4)\qquad \int_{-\pi}^{\pi} \Big| \tfrac{1}{2}\cot u/2 \int_{x}^{\pi} \tfrac{1}{2}[\cot \tfrac{1}{2}(t-u) - \cot \tfrac{1}{2}t]\,dt \Big|\, du \le \log^2 1/x + O(\log 1/x).$$

In the following calculations we shall always suppose that $\pi/2 > x > 0$. We start our proof by first integrating the inner integral of (5.4). We get

$$(5.5)\qquad \int_{x}^{\pi} \tfrac{1}{2}[\cot \tfrac{1}{2}(t-u) - \cot \tfrac{1}{2}t]\,dt = \log |\sin \tfrac{1}{2}x/\sin \tfrac{1}{2}(x-u)| - \log (1/\cos u/2).$$

Since the integral

$$\int_{-\pi}^{\pi} | \tfrac{1}{2}\cot u/2 \log 1/\cos u/2 |\, du$$

is a finite constant, it is enough to estimate

$$(5.6)\qquad I = \int_{-\pi}^{\pi} | \tfrac{1}{2}\cot u/2 \log |\sin \tfrac{1}{2}x/\sin \tfrac{1}{2}(x-u)| \,|\, du.$$

We observe that

$$\log |\sin \tfrac{1}{2}x/\sin \tfrac{1}{2}(x-u)| = \begin{cases} \log \sin \tfrac{1}{2}x/\sin \tfrac{1}{2}(u-x) & \text{for } u > x, \\ \log \sin \tfrac{1}{2}x/\sin \tfrac{1}{2}(x-u) & \text{for } u < x. \end{cases}$$

Hence I can be written as follows

$$(5.7)\qquad I = \int_{-\pi}^{x} | \tfrac{1}{2}\cot u/2 \log \sin \tfrac{1}{2}x/\sin \tfrac{1}{2}(x-u) \,|\, du + \int_{x}^{\pi} \tfrac{1}{2}\cot u/2 \,|\log \sin \tfrac{1}{2}x/\sin \tfrac{1}{2}(u-x)|\, du = I_1 + I_2,$$

say. We shall reduce I_1 to the form

$$\begin{aligned} I_1 &= \int_{0}^{x} \tfrac{1}{2}\cot u/2 \log\frac{\sin \frac{1}{2}x}{\sin \frac{1}{2}(x-u)}\,du + \int_{-\pi}^{0} -\tfrac{1}{2}\cot u/2 \,\Big|\log\frac{\sin \frac{1}{2}x}{\sin \frac{1}{2}(x-u)}\Big|\,du \\ &= \int_{0}^{x} \tfrac{1}{2}\cot u/2 \log\frac{\sin \frac{1}{2}x}{\sin \frac{1}{2}(x-u)}\,du + \int_{0}^{\pi} \tfrac{1}{2}\cot u/2 \,\Big|\log\frac{\sin \frac{1}{2}x}{\sin \frac{1}{2}(x+u)}\Big|\,du \\ &= \int_{0}^{x} \tfrac{1}{2}\cot u/2 \log\frac{\sin \frac{1}{2}x}{\sin \frac{1}{2}(x-u)}\,du + \Big(\int_{0}^{x} + \int_{x}^{2x} + \int_{2x}^{\pi}\Big) \tfrac{1}{2}\cot u/2 \log\frac{\sin \frac{1}{2}(x+}{\sin \frac{1}{2}x} \end{aligned}$$

and combine this with

$$I_2 = \int_{x}^{2x} \tfrac{1}{2}\cot u/2 \log\frac{\sin \frac{1}{2}x}{\sin \frac{1}{2}(u-x)}\,du + \int_{2x}^{\pi} \tfrac{1}{2}\cot u/2 \,\Big|\log\frac{\sin \frac{1}{2}(u-x)}{\sin \frac{1}{2}x}\Big|\,du.$$

Thus we get

$$
\begin{aligned}
I &= \int_0^x \tfrac{1}{2} \cot u/2 \log \frac{\sin \frac{1}{2}(x+u)}{\sin \frac{1}{2}(x-u)}\, du + \int_x^{2x} \tfrac{1}{2} \cot u/2 \log \frac{\sin \frac{1}{2}(x+u)}{\sin \frac{1}{2}(u-x)}\, du \\
&\quad + \int_{2x}^{\pi} \tfrac{1}{2} \cot u/2 \log \frac{\sin \frac{1}{2}(x+u) \sin \frac{1}{2}(u-x)}{\sin^2 \frac{1}{2}x}\, du \qquad (5.8) \\
&= K_1 + K_2 + K_3.
\end{aligned}
$$

Applying the transformation $\cot \frac{1}{2}x \tan \frac{1}{2}u = y$ to K_1, and noticing that

$$\frac{\sin \frac{1}{2}(x+u)}{\sin \frac{1}{2}(x-u)} = \frac{1 + \cot \frac{1}{2}x \tan \frac{1}{2}u}{1 - \cot \frac{1}{2}x \tan \frac{1}{2}u},$$

we have

$$
\begin{aligned}
(5.9) \quad K_1 &= (\cot \tfrac{1}{2}x)^2 \int_0^1 y^{-1} (y^2 + (\cot \tfrac{1}{2}x)^2)) \log ((1+y)/(1-y))\, dy \\
&\leq \int_0^1 y^{-1} \log ((1+y)/(1-y))\, dy.
\end{aligned}
$$

This integral is evidently convergent.

By applying the same transformation to K_2, we get

$$
\begin{aligned}
(5.10) \quad K_2 &\leq \int_1^{2/(1-\tan^2 x/2)} y^{-1} [\log ((y+1)/(y-1))]\, dy \\
&\leq \int_1^{\infty} y^{-1} [\log ((y+1)/(y-1))]\, dy,
\end{aligned}
$$

which is again a convergent integral.

Lastly making the change of variable $\sin \frac{1}{2}u = y \sin \frac{1}{2}x$ and noticing that $\sin \frac{1}{2}(u+x) \sin \frac{1}{2}(u-x) = (\sin \frac{1}{2}u \cos \frac{1}{2}x)^2 - (\cos \frac{1}{2}u \sin \frac{1}{2}x)^2$, we reduce K_3 to the following form

$$
\begin{aligned}
(5.11) \quad K_3 &= \int_{2\cos x/2}^{1/\sin \frac{1}{2}x} y^{-1} \log (y^2 - 1)\, dy \leq \int_{\sqrt{2}}^{1/\sin \frac{1}{2}x} y^{-1} \log y^2 dy \\
&\leq \int_1^{1/\sin \frac{1}{2}x} (2 \log y/y)\, dy = \log^2 (1/\sin \tfrac{1}{2}x) \\
&= \log^2 1/x + O(\log 1/x).
\end{aligned}
$$

Thus we have

$$I = K_1 + K_2 + K_3 \leq \log^2 1/x + O(\log 1/x).$$

Theorem 15 is thus proved.

I am grateful to Professor Zygmund for his suggestions and encouragement during the preparation of this paper.

UNIVERSITY OF CHICAGO.

REFERENCES.

1. R. Bellman, "A note on a theorem of Hardy on Fourier constants," *Bulletin of American Mathematical Society*, vol. 50 (1944), pp. 741-744.
2. G. H. Hardy, "Note on some points in the integral calculus, lxvi, The arithmetic mean of a Fourier constant," *Messenger of Mathematics*, vol. 58 (1928), pp. 50-52.
3. T. Kawata, "Notes on Fourier series (xii), On Fourier constants," *Proceedings of the Imperial Academy, Tokyo*, vol. 20 (1944), pp. 218-222.
4. G. I. Sunouchi, "On Fourier constants," *Proceedings of the Imperial Academy*, Tokyo, vol. 20 (1944), pp. 542-544.
5. A. Zygmund, "Sur les fonctions conjuguées," *Fundamenta Mathematicae*, vol. 13 (1929), pp. 284-303, Corrigenda, *ibid.*, vol. 18 (1932), p. 312.
6. ———, *Trigonometrical series*, Warszawa-Lwów (1935).

ANALYTIC CONTINUATION OF FACTORIAL SERIES.*

By V. F. Cowling.

The object of this paper is to extend to factorial series some results on the analytic continuation of Dirichlet series proved by the author.[1]

The fundamental results for factorial series are to be found in the books and papers of E. Landau [2], Milne-Thomson [3], N. Nielsen [4] and N. Nörlund [5]. A very complete and systematic study is to be found in Nörlund.

We shall prove the following theorem.

Theorem. *Let*

$$f(z) = \sum_{n=0}^{\infty} a_n n!/z(z+1)\cdot\cdot\cdot(z+n)$$

with abscissa of convergence $\gamma < +\infty$. *Let* l_1 *be a given real number. Let* $l-1 < h < l$ *where* $l > |l_1|$ *is integral and positive. Denote by* $A(h, \psi_1, \psi_2)$ *a region of the w-plane* $\psi_2 \leq \operatorname{Arg}(w-h) \leq \psi_1$ *where* $0 < \psi_1 \leq \pi/2$ *and* $-\pi/2 \leq \psi_2 < 0$.

Suppose there exists at least one set of real numbers h, l, L, K, R_1, ψ_1 *and* ψ_2, *where* h, l, ψ_1 *and* ψ_2 *are subject to the above restrictions, and a function* $a(w)$ *satisfying the following conditions simultaneously.*

a. *With the possible exception of the point at infinity* $a(w)$ *is regular in the region* $A(h, \psi_1, \psi_2)$.

b. $a(n) = a_n$, $n = l, l+1, \cdot\cdot\cdot$.

c. *For* $w = h + Re^{i\psi}$ *in the region* $A(h, \psi_1, \psi_2)$

$$a(h + Re^{i\psi}) = O[R^k \exp(-LR \sin \psi)],$$

for $R \geq R_1$ *and some* $0 < L < 2\pi$.

Then $f(z)$ *may be continued analytically over any closed bounded region contained in the half-plane* $\Re(z) > l_1$ *not containing any of the points* $z = 0, -1, -2, \cdot\cdot\cdot$.

Proof. Suppose then for a given function $f(z)$ and a given real number l_1

---

* Received May 24, 1948.

[1] See, Cowling, V. F. [1]. Numbers in brackets refer to the bibliography at the end of the paper.

there exists a set of constants h, k, L, R_1, ψ_1 and ψ_2 satisfying the conditions of the theorem.

Denote respectively by I_1, I_2 and c the sides of a contour $\Gamma(h, R')$ bounded by the lines $w = h + Re^{i\psi_1}$ and $w = h + Re^{i\psi_2}$ and arc $\psi_2 \leq \operatorname{Arg}(w - h) \leq \psi_1$ of a circle $| w - h | = R'$ of integral radius R'.

Let z be any complex number subject to the restriction that $\Re(z) > l_1$. Then since by hypothesis $l > | l_1 |$ it follows that none of the solutions in w of $w + z + 1 = N$, $N = 0, -1, -2, \cdots$ are contained in a region bounded by $\Gamma(h, R')$. Denote by B any closed bounded region contained in the half-plane $\Re(z) > l_1$ which does not contain any of the points $z = 0, -1, -2, \cdots$.

It is then simple to show by means of the Calculus of Residues that (integration being understood in the positive sense) for z contained in the half-plane $\Re(z) > l_1$

$$(1) \qquad 2\pi i \int_{\Gamma(h,R')} \frac{w\Gamma(w)\Gamma(z)a(w)\,dw}{\Gamma(z + w + 1)[\exp(2\pi iw) - 1]} = \sum_{n=l}^{R'+l-1} \frac{a(n)n!}{z(z+1)\cdots(z+n)}.$$

Set

$$(2) \qquad G(z, w) = w\Gamma(w)\Gamma(z)a(w)/\{\Gamma(z + w + 1)[\exp(2\pi iw) - 1]\}.$$

We proceed to show that

$$\int_C G(z, w)\,dw$$

converges uniformly to zero as R' becomes finite for z on a closed bounded portion of the real axis contained in the half-plane $\Re(z) > l_1$.

On the arc C

$$(3) \qquad | \exp(2\pi iw) - 1 |^{-1} = O(1).$$

This follows from the periodic character of the function and the fact that w is bounded away from an integer by our choice of h, l and R'.

From the identity

$$1/\{\exp(2\pi iw) - 1\} = \exp(-2\pi iw)/\{1 - \exp(-2\pi iw)\}$$

it follows that for w on C

$$(4) \qquad | \exp(2\pi iw) - 1 |^{-1} = O[\exp(2\pi R \sin\psi)].$$

This inequality is useful for $-\pi/2 \leq \psi < 0$.

By a well known result [4; p. 96] from the theory of the Gamma Function, if we let z denote any complex number independent of w contained in a closed bounded region, then we may write

$$\Gamma(w)w^z/\Gamma(w+z) = 1 + \epsilon(z,w) \tag{5}$$

where $\epsilon(z,w)$ tends uniformly to zero as w tends to infinity in such a way as to be bounded away from the negative real axis. In the definition of w^z we take that branch of $\log w$ which equals zero for $w = 1$.

By application of condition (c) of the hypothesis and equations (3), (4) and (5) to equation (2) we obtain after simplification, for w on c, and $z = r$ on an interval of the positive real axis contained in the half-plane $\Re(z) > l_1$,

$$\begin{aligned} (6) \quad G(z,w) &= O[R'^{k+1}\exp(-LR'\sin\psi - r\log R' - rN(R',\psi))], \\ &\qquad\qquad 0 \leq \psi \leq \pi/2; \\ &= O[R'^{k+1}\exp((2\pi - L)R'\sin\psi - r\log R' - rN(R',\psi))], \\ &\qquad\qquad -\pi/2 \leq \psi \leq 0, \end{aligned}$$

where

$$N(R',\psi) = \tfrac{1}{2}\log\left(1 + h^2/R'^2 + 2h\cos\psi/R'\right). \tag{7}$$

Clearly $N(R',\psi)$ converges uniformly to zero as R' becomes infinite for $\psi_2 \leq \psi \leq \psi_1$.

Thus after some simplification it follows from (6) that

$$\int_C G(z,w)\,dw = O[R'^{k+2-r}]. \tag{8}$$

By inspection of the right hand member of (8) it is evident that the integral along c converges uniformly to zero as R' becomes finite for $z = r$ on any closed bounded portion of the positive real axis contained in the half-plane $\Re(z) > l_1$ for which $|z| = r \geq k + 2 + \delta$, $\delta > 0$.

Hence for $z = r$ on a closed bounded portion of the positive real axis contained in the half-plane $\Re(z) > l_1$ for which $|z| = r \geqq \text{Max}\,[k + 2 + \delta, \gamma + \delta]$, $\delta > 0$ we have

$$J(z,\psi_1) + J(z,\psi_2) = \sum_{n=l}^{\infty} a(n)\,n!/z(z+1)\cdots(z+n), \tag{9}$$

where

$$J(z,\psi_1) = \int_{l_1} G(z,w)\,dw \text{ and } J(z,\psi_2) = \int_{l_2} G(z,w)\,dw.$$

We next consider under what restrictions on z, $J(z,\psi_1)$ and $J(z,\psi_2)$ converge uniformly. We omit the details in view of the fact that they are quite the same as in [1] at this stage.

By application of condition (c) of the hypothesis and equations (3) and (5) to equation (2) we find after some simplification that for $z = re^{i\theta}$ contained in a closed region B

$$(12) \qquad \int_{R_1}^{\infty} G(z, h + Re^{i\psi_1})\, dR = O\left[\int_{R_1}^{\infty} R^{k+1-\cos\theta} \exp\left(-LR \sin \psi_1\right) dR \right].$$

It follows immediately from (12) that $J(z, \psi_1)$ is uniformly convergent for $z = re^{i\theta}$ contained in the region B. Hence by a well-known theorem [6; p. 99] it follows that $J(z, \psi_1)$ represents a regular function of z in any such closed region B.

We consider next the uniform convergence of $J(z, \psi_2)$. Again it is simple to show that for z in the region B

$$(13) \qquad \int_{R_1}^{\infty} G(z, h + Rl^{i\psi_1})\, dR = O\left[\int_{R_1}^{\infty} R^{k+1-\cos\theta} \exp\left(R \sin \psi_2 (2\pi - L)\right) dR \right].$$

Since $0 < L < 2\pi$ by hypothesis and $\sin \psi_2$ is negative it is easily seen by inspection of the right hand member of (13) that $J(z, \psi_2)$ converges uniformly for $z = re^{i\theta}$ contained in a region B. As above it is then simple to demonstrate that $J(z, \psi_2)$ defines a regular function of z in a region B.

As the region B may be chosen so as to include any bounded portion of the positive real axis for which $|z| = r \geq \operatorname{Max}[k + 2 + \delta, \gamma + \delta, l_1 + \delta]$, $\delta > 0$, in its interior, it is a consequence of (9) that $J(z, \psi_1) + J(z, \psi)$ provides the analytic contitnuation of

$$f(z) - \sum_{n=l}^{\infty} a_n n!/z(z+1) \cdots (z+n)$$

over any closed region B. This completes the proof.

LEHIGH UNIVERSITY.

BIBLIOGRAPHY.

1. V. F. Cowling, "Some results for Dirichlet series," *Duke Mathematical Journal*, vol. 14, No. 4 (1947), pp. 907-912.
2. E. Landau, "Ueber die Grundlagen der Theorie der Fakultätenreihen," *Sitzungsberichte der Königlichen Bayerischen Akademie der Wissenschaften zu München*, vol. 36 (1906), pp. 151-218.
3. L. Milne-Thomson, *The Calculus of Finite Differences*, London (1933).
4. N. Nielsen, *Handbuch der Theorie der Gammafunktion*, Leipzig (1906).
5. N. Nörlund, *Leçons sur les séries d'interpolation*, Paris (1926).
6. E. Titchmarsh, *The Theory of Functions*, Oxford University Press (1939).

REPRESENTATIONS OF SEQUENCES OF SETS.*

By C. J. EVERETT [1] and G. WHAPLES.

A class of sets is representable if there exists a choice function which assigns to each set in the class an element of that set, in such a way that the same element is not assigned to two different sets. To handle in a precise manner the more general—though scarcely more difficult—case in which the same set might be counted more than once, we define a *sequence of sets* to be a mapping of a set Γ (whose elements γ shall be called indices) into subsets $M(\gamma)$ of another set E (whose members shall be called elements), with no assumption that all subsets of E appear among the $M(\gamma)$ or that different indices correspond to different subsets. We denote such a sequence by $(M(\gamma), \Gamma)$ or, if the mapping M is clear, by "sequence on Γ." By subsequences, intersections of two subsequences, etc., we mean sequences on subsets of the index set, on intersections of two subsets, etc., all with the same mapping M. A sequence is called representable if there exists a choice function f which assigns to each index γ an element $f(\gamma)$ of the corresponding set $M(\gamma)$ such that the same element is never assigned to two different indices.

1. Generalization of Hall's theorem.

A sequence shall be called *finite* if its set of indices is finite. For such sequences, P. Hall [2] has proved:

THEOREM 1. *A finite sequence $(M(\gamma), \Gamma)$ is representable if and only if it satisfies the condition*

(H) *For every integer k and every set Σ of k indices, the union of the sets $M(\sigma)$ with $\sigma \in \Sigma$ contains at least k elements.*

For the convenience of the reader we insert a new and short proof.[2] (H) is clearly necessary for representability. If we introduce the notation $\# E$ for the cardinal of a set E, and $M(\Sigma)$ for the union [3] of all sets $M(\sigma)$ with $\sigma \in \Sigma$, then (H) can be restated:

---

* Received November 5, 1946; revised May 21, 1948. Presented to the American Mathematical Society, April 25, 1947.

[1] Part of the work of this author was supported by the University of Wisconsin Alumni Research Foundation.

[2] The basic idea of this proof was suggested by a conversation with P. Erdös; he has an independent proof of our Theorem 2, using transfinite induction and the above method of deletions. W. Gustin also contributed improvements.

[3] We shall use this abbreviation throughout the paper. A more common convention would have $M(\Sigma)$ mean the class of all sets $M(\sigma)$, rather than their union; but we have no need of this class of sets.

$$\# M(\Sigma) \geq \# \Sigma \textit{ for every } \Sigma \subseteq \Gamma.$$

Call the subsequence on Σ *perfect* if $\# M(\Sigma) = \# \Sigma$.

LEMMA 1. *The union and intersection of any two perfect subsequences of a sequence which satisfies* (H) *are perfect.*

Proof. Let the sequences on Σ_1 and Σ_2 be perfect. By (H) we have already:

$$\# M(\Sigma_1 \cup \Sigma_2) \geq \# (\Sigma_1 \cup \Sigma_2).$$

Also,

$$\# (\Sigma_1 \cup \Sigma_2) = \# \Sigma_1 + \# \Sigma_2 - \# (\Sigma_1 \cap \Sigma_2)$$

and

$$\begin{aligned} \# M(\Sigma_1 \cup \Sigma_2) &\leq \# M(\Sigma_1) + \# M(\Sigma_2) - \# M(\Sigma_1 \cap \Sigma_2) \\ &= \# \Sigma_1 + \# \Sigma_2 - \# M(\Sigma_1 \cap \Sigma_2). \end{aligned}$$

Since, by (H), $\# M(\Sigma_1 \cap \Sigma_2) \geq \# (\Sigma_1 \cap \Sigma_2)$, we get $\# M(\Sigma_1 \cup \Sigma_2) \leq \# (\Sigma_1 \cup \Sigma_2)$, and the lemma follows.

Proof of theorem. It is trivial for all sequences with $\# \Gamma = 1$; assume it for all sequences with less than n indices and suppose that $(M(\gamma), \Gamma)$ satisfies (H) and that $\Gamma = \{1, 2, \cdots, n\}$. For each element e of the set $M(n)$, consider the sequence obtained by deleting e (if it occurs) from each $M(\gamma)$ and letting γ run over $\Gamma' = \{1, 2, \cdots, (n-1)\}$. If, for any such e, the corresponding sequence should satisfy (H), then it would be representable and we could get a representation of the original sequence by assigning to n this element e. Suppose then that none of these new sequences satisfies (H). Then there must exist, for each $e \varepsilon M(n)$, a set of indices Σ_e such that:

$$\Sigma_e \subseteq \Gamma', \text{ hence } n \notin \Sigma_e, \text{ and } \# (M(\Sigma_e) - \{e\}) < \# \Sigma_e.$$

Since deletion of e can change the cardinal of $M(\Sigma_e)$ by at most 1, and since the original sequence satisfied (H), we see from the second relation that $M(\Sigma_e)$ always contains e and that the sequences $(M(\sigma), \Sigma_e)$ are perfect. Let Σ' be the union of the Σ_e; then the sequence on Σ' is perfect and $M(\Sigma') \supset M(n)$. Hence the set Σ, obtained by adjoining n to the set Σ', contradicts (H). $(\# M(\Sigma) = \# M(\Sigma') = \# \Sigma - 1.)$

Hall's theorem would be false if the restriction to finite index sets were dropped. To see this, let Γ be the set of all positive integers together with the symbol ω, and define $M(n) = \{n\}$, $M(\omega) =$ set of all positive integers. Then (H) is true, even for infinite index sets. But one sees easily that no number will work as the representative for ω, so this sequence is not representable. In view of our Theorem 2′, we point out that it has the further property that every finite subsequence is representable.

We prove however the following generalization of Hall's theorem, which has no restriction on the cardinal of the index set:[4]

THEOREM 2. *If* $(M(\gamma), \Gamma)$ *is any sequence of finite sets (i. e. if each* $M(\gamma)$ *is finite) then it is representable if and only if* $\# M(\Sigma) \geqq \# \Sigma$ *for every finite subset* Σ *of* Γ.

In view of Hall's result, this is equivalent to:

THEOREM 2′. *A sequence of finite sets is representable if and only if every finite subsequence is representable.*

Proof. Let $(M(\gamma), \Gamma)$ be a sequence of finite sets, with every finite subsequence representable. A function f, defined on a subset Δ of the index set to the set E, shall be called a *universal representation function* (u. r. f.) if every finite subsequence has a representation which agrees with f for all indices for which f is defined—i. e. if each $(M(\sigma), \Sigma)$ with finite Σ has a representation r with $r(\sigma) = f(\sigma)$ on $\Sigma \cap \Delta$. (This implies, of course, that $f(\delta) \in M(\delta)$—consider the subsequence on the single index δ.)

If $\emptyset$ denotes the null set, and $*$ the "function whose range of definition is $\emptyset$," then $*$ shall be considered as a function, and hence a u. r. f. The fact that it is one is not a complete triviality, since it demands that every finite subsequence be representable—viz., the main assumption of the theorem.

We prove our theorem by showing that there is a u. r. f. defined over the whole set Γ. Our proof uses the lemma of Zorn: *Any non-empty partially ordered set, in which each linearly ordered subset has an upper bound, contains a maximal element* [9]. We partially order the set of u. r. f. by defining one of them to be less than a second if the range of definition of the second includes that of the first, and if they agree whenever both are defined.

LEMMA 2. *Every linearly ordered set of u. r. f. has an upper bound.*

(This lemma needs no assumption about the sequence.) Let $\mathfrak{L}$ be a linearly ordered set of u. r. f., Λ the set of all indices for which any one of them is defined: we must show that there is a u. r. f. defined on all of Λ and agreeing with every function in $\mathfrak{L}$. How to define this function is clear:

---

[4] R. Rado, [4], a paper which was called to our attention by the referee, gives an entirely different generalization. He keeps the index set finite (except for one unproved remark) but replaces the cardinal numbers of the sets by a general metric, and extends the idea of a representation.

if $\lambda \varepsilon \Lambda$, define $l(\lambda)$ to be $f(\lambda)$, where f is any function in $\mathcal{L}$ which is defined for λ. One sees that there is at least one such f, and that $l(\lambda)$ does not depend on which one we choose. If Σ is any finite index set, there is some f in $\mathcal{L}$ which is defined for all indices in Σ: since f is a u. r. f., Σ has a representation which agrees with f and hence also with l.

LEMMA 3. *Let f be a u. r. f., with range of definition Δ, for a sequence $(M(\gamma), \Gamma)$ of finite sets; and let τ be any index not in Δ. Then f can be extended to a u. r. f whose range of definition is $\Delta \bigcup \{\tau\}$.*

Suppose the lemma false. Then for every element e of the finite set $M(\tau)$, the extension of f obtained by assigning to τ the representative e fails to be a u. r. f. Then there exists for each $e \varepsilon M(\tau)$ a finite set of indices Σ_e, such that $\tau \varepsilon \Sigma_e$ and every representation of the sequence on Σ_e either disagrees with f or assigns to τ a representative different from e. For each $e \varepsilon M(\tau)$ choose one such Σ_e, and let Σ be their union. Since $M(\tau)$ was finite, Σ is finite. Any representation of $(M(\sigma), \Sigma)$ would assign to τ some representative $e \varepsilon M(\tau)$, would induce a representation of the subsequence on Σ_e, and hence would disagree with f. This contradicts the assumption that f was a u. r. f., and proves the lemma.

These two lemmas, together with the lemma of Zorn, clearly imply Theorem 2. Indeed, if we consider the set of all u. r. f. which exceed a given one, we get the somewhat stronger proposition: *Any u. r. f. of a sequence of finite sets can be extended to a representation of the whole sequence.*

2. Representations relative to a partition. A partition of a set E is a class $\mathfrak{P} = \{\mathfrak{p}\}$ of disjoint subsets of E, whose union is the whole set. (Nature of our proofs will justify our choice of notation.) A sequence $(M(\gamma), \Gamma)$ of subsets of E has a representation *relative to* $\mathfrak{P}$ (abbreviated to $\mathfrak{P}$-representation) if there is a function f which, in addition to being a representation, satisfies the condition that representatives of different indices are contained in different elements of $\mathfrak{P}$ ($\alpha \neq \beta$, $f(\alpha) \varepsilon \mathfrak{a}$, $f(\beta) \varepsilon \mathfrak{b}$ implies $\mathfrak{a} \neq \mathfrak{b}$).

A subset of E shall be called finite with respect to $\mathfrak{P}$ ($\mathfrak{P}$-finite) when it meets only a finite number of elements of $\mathfrak{P}$. A class or sequence of sets is $\mathfrak{P}$-finite when all its members are.

THEOREM 3. *If $\mathfrak{P}$ is a partition of E, and if $(M(\gamma), \Gamma)$ is a $\mathfrak{P}$-finite sequence of subsets of E, this sequence has a $\mathfrak{P}$-representation if and only if, for every finite index set Σ, the number of $\mathfrak{p} \varepsilon \mathfrak{P}$ which meet $M(\Sigma)$ is at least equal to the number of indices in Σ.*

Proof. (Method due to P. Hall [2]). For each $\gamma \in \Gamma$, define the subset $\mathfrak{M}(\gamma)$ of $\mathfrak{P}$ to be the set of all $\mathfrak{p} \in \mathfrak{P}$ which contain elements of $M(\gamma)$. Since each $M(\gamma)$ is $\mathfrak{P}$-finite, the $\mathfrak{M}(\gamma)$ are finite sets. The number of elements (of $\mathfrak{P}$!) in $\mathfrak{M}(\mathfrak{S})$ is at least $\# \mathfrak{S}$, for each finite $\mathfrak{S}$. So by Theorem 2, sequence $(\mathfrak{M}(\gamma), \Gamma)$ has a representation $\mathfrak{f}$, mapping Γ into $\mathfrak{P}$. Any f which maps Γ into E in such a way that $f(\gamma) \in \mathfrak{f}(\gamma)$ for all γ (axiom of choice) is a $\mathfrak{P}$-representation.

3. de Bruijn's theorem.

Suppose E has two partitions, $\mathfrak{A} = \{\mathfrak{a}\}$ and $\mathfrak{B} = \{\mathfrak{b}\}$. We shall say that $\mathfrak{A}$ and $\mathfrak{B}$ have a common representation when there exists a one-one correspondence $\mathfrak{a} \sim \mathfrak{b}$ between $\mathfrak{A}$ and $\mathfrak{B}$ for which corresponding $\mathfrak{a}$'s and $\mathfrak{b}$'s have non-empty intersections, together with functions f defined on $\mathfrak{A}$ and g on $\mathfrak{B}$ such that $f(\mathfrak{a}) = g(\mathfrak{b}) \in \mathfrak{a} \cap \mathfrak{b}$ whenever $\mathfrak{a} \sim \mathfrak{b}$.

We can prove very easily the following theorem, which was first proved by de Bruijn (although our work was independent of his).

THEOREM 4. *If $\mathfrak{A}$ and $\mathfrak{B}$ are two partitions of a set E, and if each of them is finite with respect to the other, then they have a common representation if and only if, for all integers k, the union of each k elements $\mathfrak{a} \in \mathfrak{A}$ meets at least k elements of $\mathfrak{B}$, and also the union of each k elements $\mathfrak{b} \in \mathfrak{B}$ meets at least k elements of $\mathfrak{A}$.*

Proof. Apply Theorem 2, with $\mathfrak{A}$ as the class of sets and $\mathfrak{B}$ the partition: $\mathfrak{A}$ has a representation relative to $\mathfrak{B}$. Similarly, $\mathfrak{B}$ has a representation relative to $\mathfrak{A}$. We only need further:

LEMMA 4. *If $\mathfrak{A}$ and $\mathfrak{B}$ are partitions of a set, if $\mathfrak{A}$ has a representation relative to $\mathfrak{B}$, and if $\mathfrak{B}$ has a representation relative to $\mathfrak{A}$, then $\mathfrak{A}$ and $\mathfrak{B}$ have a common representation.*

Proof. Let f be the representation of $\mathfrak{A}$ relative to $\mathfrak{B}$. Then if $\mathfrak{f}(\mathfrak{a})$ denotes that element of partition $\mathfrak{B}$ which contains $f(\mathfrak{a})$, this function $\mathfrak{f}$ defines a one-one correspondence between $\mathfrak{A}$ and a subset of $\mathfrak{B}$. In just the same way we get a one-one function $\mathfrak{g}$ mapping $\mathfrak{B}$ onto a subset of $\mathfrak{A}$. A theorem of Banach [5, p. 90] asserts that in just this situation there exist disjoint splittings $\mathfrak{A} = \mathfrak{A}_1 \cup \mathfrak{A}_2$ and $\mathfrak{B} = \mathfrak{B}_1 \cap \mathfrak{B}_2$ such that $\mathfrak{f}$ gives a one-one correspondence between $\mathfrak{A}_1$ and $\mathfrak{B}_1$ and $\mathfrak{g}$ a one-one correspondence between $\mathfrak{B}_2$ and $\mathfrak{A}_2$. This correspondence, together with the function h, defined to be f on $\mathfrak{A}_1$ and $g\mathfrak{g}^{-1}$ on $\mathfrak{A}_2$, gives the desired representation.

COROLLARY 4. 1. *If* $\mathfrak{A}$ *and* $\mathfrak{B}$ *are two partitions of* E *for which all* $\mathfrak{a}$*'s and all* $\mathfrak{b}$*'s have the same finite number* h *of elements, then* $\mathfrak{A}$ *and* $\mathfrak{B}$ *have a common representation.*

For the union of k members of $\mathfrak{A}$ contains kh elements of E, hence meets at least k elements of $\mathfrak{B}$.

COROLLARY 4. 2. *The system of left and right cosets of a finite subgroup of any group have a common representation.*

König and Valkó [3] proved 4. 1. Van der Waerden [7] used their result to establish 4. 2, and showed that this corollary can be false for an infinite subgroup of infinite index, but proved it (using group theory and Theorem 4 applied to finite partitions) for all subgroups of finite index. A proof of this case of the corollary can also be found in Zassenhaus, [8, pp. 11 and 35].

4. Concluding remarks. The full power of Theorem 1 is not needed to prove Theorem 4; de Bruijn uses the following trick, similar to one of König, for reducing this theorem to the denumerable case. Call $\mathfrak{a} \varepsilon \mathfrak{A}$ and $\mathfrak{b} \varepsilon \mathfrak{B}$ *immediately connected* if their intersection is non-null, and any two elements of $\mathfrak{A} \cup \mathfrak{B}$ *connected* if there exists a finite chain of immediately connected elements leading from one to the other: one sees at once that $\mathfrak{A}$ and $\mathfrak{B}$ fall into denumerable, connected components and that if $\mathfrak{A}'$, $\mathfrak{B}'$ are components of $\mathfrak{A}$ and $\mathfrak{B}$, then the sets $\bigcup_{\mathfrak{A}'}\mathfrak{a}$ and $\bigcup_{\mathfrak{B}'}\mathfrak{b}$ are either disjoint or equal.

So only the denumerable case of Theorem 2 is needed for de Bruijn's Theorem. One may then avoid the theorem of Banach, also, by using an alternating process to build the common representation so as to insure that each $\mathfrak{a} \varepsilon \mathfrak{A}$ and each $\mathfrak{b} \varepsilon \mathfrak{B}$ has a representative, using Lemma 3 at each step.

No similar reduction of Theorem 1 to the denumerable case seems possible: and we would conjecture that the axiom of choice (or lemma of Zorn) for arbitrary cardinals is really necessary for proof of this theorem. Certainly no such simple decomposition as de Bruijn uses is available in this case. For example, let $(x, y) \rightarrow z$ be a one-one correspondence of points of the unit square to points of unit interval, and consider the class of sets (of one, two, or three elements) $\{x, y, z\}$ where $(xy) \rightarrow z$. It is nondenumerable and has a representation (take z as the representative), but every set is connected with every other set.

Theorem 2 can be proved much more simply in the denumerable case. Let Γ be the set of positive integers, suppose $(M(\gamma), \Gamma)$ satisfies the assumptions of Theorem 2, and consider all functions f, defined on Γ, with $f(\nu) \varepsilon M(\nu)$ for each $\nu \varepsilon \Gamma$. Let Δ_n be the set of all such functions for which the values $f(1), f(2), \cdots, f(n)$ are all distinct. We have to show that the intersection of all the sets Δ_n is not empty. If we make the set of all these functions f into a metric space by defining $\rho(f, g) = 1/n$ where n is the smallest integer ν for which $f(\nu) \neq g(\nu)$ [6, p. 74], then the Δ_n are closed and compact, hence have a non-null intersection by the Cantor theorem [6, p. 30]. One could also show this fact by use of the Cantor diagonal process.

UNIVERSITY OF WISCONSIN AND INDIANA UNIVERSITY.

BIBLIOGRAPHY.

1. N. G. de Bruijn, "Gemeenschappelijke representantensystemen van twee klassenindeelingen van een verzameling," *Niew Archiev Wiskunde* (2), vol. 22 (1943), pp. 48-52.
2. P. Hall, "On representation of subsets," *Journal of the London Mathematical Society*, vol. 10 (1935), pp. 26-30.
3. D. König and S. Valkó, "Über mehrdeutige Abbildungen von Mengen," *Mathematische Annalen*, vol. 95 (1925), pp. 135-138.
4. R. Rado, "A theorem on general measure functions," *Proceedings of the London Mathematical Society*, vol. 44 (1938), pp. 61-91.
5. W. Sierpinski, *Leçons sur les nombres transfinis*, Paris, 1928.
6. ———, *Introduction to General Topology*, University of Toronto Press, 1934.
7. B. L. van der Waerden, "Ein Satz über Klasseneinteilungen von endlichen Mengen," *Hamburger Abhandlungen*, vol. 5 (1927), pp. 185-188.
8. H. Zassenhaus, *Lehrbuch der Gruppentheorie*, Berlin, 1937.
9. M. Zorn, "A remark on method in transfinite algebra," *Bulletin of the American Mathematical Society*, vol. 41 (1935), pp. 667-670.

METRIC PROPERTIES OF DIFFERENTIAL EQUATIONS.*

By D. C. Lewis.

1. Introduction.

The purpose of this paper is to discuss the dependence of solutions of a system of differential equations,

$$(1.1) \qquad dx_i/dt = X^i(x_1, \cdots, x_n, t) = X^i[x, t], \qquad i = 1, 2, \cdots, n,$$

on the initial conditions. The classical existence proofs yield inequalities which show that the solutions depend continuously on the initial conditions, at least, if the right hand members satisfy a Lipschitz condition. In this paper we shall refine and supplement these classical inequalities, assuming that the X^i are of class C'. In so doing we call attention to the fundamental importance of a certain quadratic form Q to be introduced in the sequel. Still further, but less useful, refinements may be obtained on the assumption that the X^i are of class C'', in which case a certain biquadratic form B plays the decisive role. It is also indicated, by consideration of simple examples, that our fundamental theorems are the best possible ones of their type. A simple application to the theory of qualitative integration is given.

Our theorems are specified in terms of a metric which we assign to the space of the $(x_1, \cdots, x_n)$. For most purposes it is sufficient to assume a Riemannian space. Much of the analysis, however, is not more difficult, if we assume a general Finsler space.

Let $f[x, \dot{x}] = f(x_1, \cdots, x_n, \dot{x}_1, \cdots, \dot{x}_n)$ be of class C''' and positively homogeneous of degree $+1$ in $\dot{x}_1, \cdots, \dot{x}_n$. Furthermore, suppose $f[x, \dot{x}] > 0$, unless all the $\dot{x}$'s are zero, and that $f_{\dot{x}_i \dot{x}_j}[x, \dot{x}]\lambda^i\lambda^j > 0$ for all sets $\lambda^1, \cdots, \lambda^n$, not proportional to $\dot{x}_1, \cdots, \dot{x}_n$.[1] Here, as in the sequel, the summation convention of tensor analysis is used, with all repeated indices summed from 1 to n. The length of an arc, $x_i = x_i(\tau)$, $i = 1, \cdots, n$, $0 \leqq \tau \leqq 1$, is defined as the value of the integral, $\int_0^1 f[x(\tau), x'(\tau)]d\tau$, which exists if the arc is sufficiently regular and is, moreover, independent of the parametrization on

---

* Received May 11, 1948.

[1] The purpose of this condition is to insure the "regularity" of the variational problem $\delta \int f[x, \dot{x}]d\tau = 0$ in parametric form. Cf., for instance, Marston Morse, "The calculus of variations in the large," *American Mathematical Society Colloquium Publications*, vol. 18 (1934), p. 121.

account of the assumed homogeneity of $f[x, \dot{x}]$. The *distance* between two points is the length of the shortest arc connecting the two points. The parametric equations of such a *geodesical* arc are well known to satisfy Euler's equations,

$$f_{x_i}[x(\tau), x'(\tau)] - (d/d\tau) f_{\dot{x}_i}[x(\tau), x'(\tau)] = 0, \qquad i = 1, \cdots, n. \tag{1.2}$$

2. Fundamental results on the hypothesis that the X^i are of class C'.

THEOREM 1. *Let us consider two integral arcs of the system* (1.1).

$$C_0\colon\ x_i = x_i(t, 0) \ \textit{ and } \ C_1\colon\ x_i = x_i(t, 1), \qquad t_0 \leqq t \leqq t_0 + h.$$

We let $D(t)$ denote the distance between $[x(t, 0)]$ and $[x(t, 1)]$ measured along the geodesical arc g_t. We furthermore assume that all these g_t are included within an open region R, and that no complete geodesic obtained by extending a g_t contains entirely within R an arc whose end points are conjugate to each other (in the sense of the calculus of variations).

If

$$\alpha(t) \leqq \{f_{x_i}[x, \lambda] X^i[x, t] + f_{\dot{x}_i}[x, \lambda] X^i{}_{x_j}[x, t] \lambda^j\} \leqq \beta(t) \tag{2.1}$$

for every set $\lambda^1, \cdots, \lambda^n$ such that $f[x, \lambda] = 1$ and at every point $[x]$ on g_t, then

$$D(t_0) \exp \int_{t_0}^{t} \alpha(u)\, du \leqq D(t) \leqq D(t_0) \exp \int_{t_0}^{t} \beta(u)\, du, \qquad t_0 \leqq t \leqq t_0 + h. \tag{2.2}$$

Before proceeding with the proof of this theorem we note that in the important Riemannian case, when

$$f[x, \dot{x}] = (g_{ij}[x] \dot{x}^i \dot{x}^j)^{\frac{1}{2}}, \tag{2.3}$$

the middle member of (2.1) can be replaced by a very simple quadratic form in $\lambda', \cdots, \lambda^n$, namely,

$$Q[\lambda] = X_{ij} \lambda^i \lambda^j = (\tfrac{1}{2} X^k \partial g_{ij}/\partial x_k + g_{ik} X^k{}_{x_j}) \lambda^i \lambda^j. \tag{2.4}$$

Here the λ's are components of a unit vector, inasmuch as $f[x, \lambda] = 1$ can now be written in the form $g_{ij} \lambda^i \lambda^j = 1$. In formula (2.4), X_{ij} is the covariant derivative of the covariant vector $X_i = g_{ik} X^k$.

To prove the theorem, we let

$$x_i = x_i(t, \tau), \qquad i = 1, \cdots, n, \ 0 < \tau < 1, \tag{2.5}$$

be the coördinates of the point on g_t which divides g_t into two arcs of shorter length in the ratio $\tau:(1-\tau)$ from $[x(t,0)]$ to $[x(t,1)]$. The functions $x_i(t,\tau)$ may be shown to be of class C''. It is here that the hypothesis concerning the non-existence of certain pairs of conjugate points in R is used. By definition of $D(t)$ we have

$$D(t)=\int_0^1 f[x(t,\tau),x_\tau(t,\tau)]\,d\tau, \tag{2.6}$$

which we differentiate under the integral sign, thus obtaining

$$D'(t)=\int_0^1 (f_{x_i}[x,x_\tau](\partial x_i/\partial\tau)+f_{\dot{x}_i}[x,x_\tau](\partial^2x_i/\partial t\partial\tau))\,d\tau.$$

The second term in the integrand can be integrated by parts, and, in a manner familiar to all devotees of the calculus of variations, we find that

$$\begin{aligned}D'(t)=&\int_0^1\{(f_{x_i}[x,x_\tau]-(\partial/\partial\tau)f_{\dot{x}_i}[x,x_\tau])\partial x_i/\partial\tau\}d\tau\\&+\{f_{\dot{x}_i}[x(t,1),x_\tau(t,1)]\partial x_i(t,1)/\partial t-f_{\dot{x}_i}[x(t,0),x_\tau(t,0)]\partial x_i(t,0)/\partial t\}.\end{aligned}$$

In virtue of (1.2), which the $x_i(t,\tau)$ considered as functions of τ are known to satisfy, the integral in the above representation of $D'(t)$ disappears at once. Remembering that the curves C_0 and C_1 are integral curves of (1.1), the rest of the expression can also be simplified, with the result that

$$\begin{aligned}D'(t)=&\{f_{\dot{x}_i}[x(t,1),x_\tau(t,1)]X^i[x(t,1),t]\\&-f_{\dot{x}_i}[x(t,0),x_\tau(t,0)]X^i[x(t,0),t]\}\\=&J(t,1)-J(t,0),\end{aligned} \tag{2.7}$$

where we define J as follows:

$$J(t,\tau)=f_{\dot{x}_i}[x(t,\tau),x_\tau(t,\tau)]X^i[x(t,\tau),t].$$

Differentiating this last equality and again using (1.2), we obtain

$$\begin{aligned}J_\tau(t,\tau)=&f_{x_i}[x(t,\tau),x_\tau(t,\tau)]X^i[x(t,\tau),t]\\&+f_{\dot{x}_i}[x(t,\tau),x_\tau(t,\tau)]X^i{}_{x_j}[x(t,\tau),t]\partial x_j/\partial\tau.\end{aligned} \tag{2.8}$$

We now introduce the new variable s defined by

$$s=s(t,\tau)=\int_0^\tau f[x(t,\tau'),x_\tau(t,\tau')]\,d\tau' \tag{2.9}$$

which has the obvious properties:

$$\partial s/\partial\tau = f[x(t,\tau), x_\tau(t,\tau)], \tag{2.10}$$

$$s(t,1) = D(t) \tag{2.11}$$

$$f[x, \partial x/\partial s] = 1. \tag{2.12}$$

The last of these relations follows from the fact that, on account of the assumed homogeneity of $f[x,\dot{x}]$, the parameter in (2.9) can be changed from τ' to s', where $s' = s(t,\tau)$, with the result, deduced from (2.9), that

$$s = \int_0^s f[x(t,s'), x_s(t,s')]ds'.$$

Again using the homogeneity property of $f[x,\dot{x}]$ and the relations $x_\tau = x_s(\partial s/\partial\tau] = x_s f[x, x_\tau]$, we obtain from (2.8) the result that

$$J_\tau(t,\tau) = [f_{x_i}[x(t,s), x_s]X^i[x(t,s),t] \\ + f_{\dot{x}_i}[x, x_s]\,(\partial X^i/\partial x_j)\,(\partial x_j/\partial s)]f[x, x_\tau].$$

From this we find with the aid of (2.1) and (2.12) that

$$\alpha(t)f[x, x_\tau] \leqq J_\tau(t,\tau) \leqq \beta(t)f[x, x_\tau].$$

Integrating from $\tau = 0$ to $\tau = 1$ and referring to (2.6) and (2.7), we find that

$$\alpha(t)D(t) \leqq D'(t) \leqq \beta(t)D(t). \tag{2.13}$$

The proof of the theorem is completed by integration of these last two inequalities.

The theorem just proved yields at once a result more reminiscent of the classical inequalities, if we take for $\alpha(t)$ and $\beta(t)$ constants $\bar{\alpha}$ and $\bar{\beta}$ which are respectively lower and upper bounds in R for the middle member of (2.1) or (in the case of a Riemannian metric) for the quadratic form Q given by (2.4). We thus obtain the result

$$D(t_0)e^{\bar{\alpha}(t-t_0)} \leqq D(t) \leqq D(t_0)e^{\bar{\beta}(t-t_0)}, \qquad t_0 \leqq t, \tag{2.14}$$

which may be compared with such a classical formula[2] as

$$D(t) \leqq D(t_0)e^{M(t-t_0)}. \tag{2.15}$$

---

[2] Cf., for instance, formula (2), p. 43, of Bieberbach's textbook, *Differentialgleichungen*, where a different notation is used in connection with a first order system and the usual Euclidean metric.

Here M is the positive constant appearing in the Lipschitz condition (or some positive constant multiple thereof in the more complicated cases). In other words, if the X's are of class C' and if we neglect a possible positive constant factor, M is an upper bound for the $|\partial X^i/\partial x_j|$. As we shall show in **6**, there are cases when Q is negative definite and hence $\bar{\beta}$ may sometimes be taken to be zero or even negative, whereas (except for the trivial case when the X's are independent of $x_1, \cdots, x_n$), M must be positive. In this sense, at least, (2.14) is an essential refinement of (2.15).

THEOREM 2. *Let* C_0 *and* C_1 *be defined as in the preceding theorem. Let* Γ_{t_0} *be an arbitrary sufficiently regular curve, with parametric equations,* $x_i = \xi_i(\tau)$, $i = 1, \cdots, n$, $0 \leqq \tau \leqq 1$, *whose end points are at* $[x(t_0, 0)]$ *and* $[x(t_0, 1)]$ *respectively. Let* C_τ *be the curve whose parametric equations,* $x_i = x_i(t, \tau)$, *satisfy the system* (1.1) *for each fixed value of* τ *on the unit interval and take on the initial values* $x_i(t_0, \tau) = \xi_i(\tau)$. *Let* Γ_t *be the arc whose parametric equations, for each fixed* t, $(t_0 \leqq t \leqq t_0 + h)$, *are likewise* $x_i = x_i(t, \tau)$, $0 \leqq \tau \leqq 1$. *Let* $L(t)$ *denote the length of* Γ_t.

If

$$\alpha^*(t) \leqq \{f_{x_i}[x, \lambda] X^i[x, t] + f_{\dot{x}_i}[x, \lambda] X^i{}_{x_j}[x, t] \lambda^j\} \leqq \beta^*(t), \tag{2.16}$$

for every set of the λ'*s, such that* $f[x, \lambda] = 1$ *and at every point* $[x]$ *of* Γ_t, *then*

$$L(t_0) \exp \int_{t_0}^{t} \alpha^*(u)\, du \leqq L(t) \leqq L(t_0) \exp \int_{t_0}^{t} \beta^*(u)\, du. \tag{2.17}$$

To prove this, we observe that by our definition of length

$$L(t) = \int_0^1 f[x(t, \tau), x_\tau(t, \tau)]\, d\tau. \tag{2.18}$$

Differentiating under the integral sign, we have (after making use of (1.1))

$$L'(t) = \int_0^1 \{f_{x_i}[x(t, \tau), x_\tau(t, \tau)] X^i[x(t, \tau), t] + f_{\dot{x}_i}[x(t, \tau), x_\tau(t, \tau)] X^i{}_{x_j}[x(t, \tau), t] \partial x_j/\partial \tau\}\, d\tau. \tag{2.19}$$

As in the preceding proof, we introduce the arc length s as a new parameter.

$$s = \int_0^\tau f[x(t, \tau'), x_\tau(t, \tau')]\, d\tau', \qquad f[x, \partial x/\partial s] = 1.$$

Again using the homogeneity of $f[x, \dot{x}]$ and hence the same homogeneity of the integrand in (2.19), we find that

$$(2.20)\qquad L'(t)=\int_0^{L(t)}\{f_{x_i}[x,\partial x/\partial s]X^i[x,t]+f_{\dot{x}_i}[x,\partial x/\partial s]X^i{}_{x_j}[x,t]\partial x_j/\partial s\}ds.$$

It now follows at once from (2.16) that

$$(2.21)\qquad \alpha^*(t)L(t)\leqq L'(t)\leqq \beta^*(t)L(t),$$

and hence the proof is completed by integration of (2.21).

The relationship between Theorem 1 and Theorem 2 is made sufficiently obvious by the following remark:

Although Γ_t is not, in general, a geodesic for all values of t (or even for any value of t), it is possible to choose Γ_{t_0} in such a way that for some particular t, Γ_t is a geodesic. For such a Γ_t it is clear that

$$D(t)=L(t)$$

whereas

$$D(t+\Delta t)\leqq L(t+\Delta t).$$

Hence

$$\{D(t+\Delta t)-D(t)\}/\Delta t\leqq\{L(t+\Delta t)-L(t)\}/\Delta t,\ \text{if}\ \Delta t>0,$$

while

$$\{D(t+\Delta t)-D(t)\}/\Delta t\geqq\{L(t+\Delta t)-L(t)\}/\Delta t,\ \text{if}\ \Delta t<0.$$

Hence, allowing Δt to approach zero first through positive values and then through negative values, we find that $D'(t)=L'(t)$. Thus the inequalities (2.13) used in the proof of Theorem 1 can be deduced from the corresponding inequalities (2.21) used in the proof of Theorem 2. The indicated alternative in the proof Theorem 1 was not adopted because of the essential importance of formula (2.7) in 3.

THEOREM 3. *Let C_0, C_1, $D(t)$, $\alpha(t)$ be defined as in Theorem* 1. *Let the Γ_{t_0} introduced in Theorem* 2 *be chosen as a geodesic, and then let $L(t)$ and $\beta^*(t)$ be defined as in Theorem* 2. *Then*

$$0\leqq L(t)-D(t)\leqq D(t_0)[\exp\int_{t_0}^t\beta^*(u)du-\exp\int_{t_0}^t\alpha(u)du],$$
$$t_0\leqq t\leqq t_0+h.$$

This theorem, which is a trivial consequence of $L(t_0)=D(t_0)$ and the two previous theorems, gives some information about the manner in which geodesics are deformed under the transformations defined by the differential system (1.1).

3. Estimates for $D''(t)$ and $L''(t)$ on the hypothesis that the X^i are of class C''. So far we have not made use of the particular manner in which the parameter τ was chosen in the proof of Theorem 1. Actually it was chosen in such a manner that

$$s(t,\tau) = \tau D(t) = \int_0^\tau f[x(t,\tau'), x_\tau(t,\tau')]d\tau',$$

whence, differentiating with respect to τ, we have

$$D(t) = f[x(t,\tau), x_\tau(t,\tau)] \tag{3.1}$$

$$0 = f_{x_i}(\partial x_i/\partial\tau) + f_{\dot{x}_i}(\partial^2 x_i/\partial\tau^2). \tag{3.2}$$

On account of the assumptions made in **1** on $f[x, \dot{x}]$, it may be shown that, although the n equations (1.2) are not independent, these equations taken together with (3.2) can be solved uniquely[3] for $\partial^2 x_i/\partial\tau^2$ in terms $x_1, \cdots, x_n$ and $\partial x_1/\partial\tau, \cdots, \partial x_n/\partial\tau$. We accordingly write

$$\partial^2 x_i/\partial\tau^2 = F^i[x, \partial x/\partial\tau]. \tag{3.3}$$

It is easy to verify that the functions $F^i[x, \dot{x}]$, here introduced, must be homogeneous of degree 2 in $\dot{x}_1, \cdots, \dot{x}_n$. It is perhaps also worthwhile to record the fact that these F's are determined by the following system of n linear equations:

$$(ff_{\dot{x}_i\dot{x}_j} + f_{\dot{x}_i}f_{\dot{x}_j})F^j = ff_{x_i} - (ff_{\dot{x}_i x_k} + f_{\dot{x}_i}f_{x_k})\dot{x}_k, \qquad i = 1, \cdots, n. \tag{3.4}$$

We introduce the abbreviations,

$$\begin{aligned} P[t, x, \dot{x}] = {} & f_{\dot{x}_i\dot{x}_j}\{X^iX^j{}_{x_k}F^k - X^iX^kF^j{}_{x_k} - \tfrac{1}{2}(X^iX^k)_{x_l}\dot{x}_l F^j{}_{\dot{x}_k}\} \\ & + f_{\dot{x}_i\dot{x}_j}(X^iX^j{}_{x_k})_{x_h}\dot{x}_h\dot{x}_k + f_{x_i\dot{x}_j}(X^iX^j)_{x_k}\dot{x}_k + f_{x_ix_j}X^iX^j \\ & + f_{\dot{x}_i}\{(X^i{}_{x_j}X^j)_{x_k} + X^i{}_{tx_k}\}\dot{x}_k \\ & + f_{x_i}\{X^i{}_{x_j}X^j + X^i{}_t\}, \end{aligned} \tag{3.5}$$

---

[3] These are essentially well known facts in the Calculus of Variations. It is also easily shown that a set of n independent equations equivalent to the $n+1$ equations (1.2) and (3.2) may be written

$$[\partial\Phi(f)/\partial x_i] - [\partial/\partial\tau][\partial\Phi(f)/\partial\dot{x}_i] = 0,$$

where Φ is an arbitrary function of class C'' whose first and second derivatives are positive for positive values of its argument. Formula (3.4) was obtained by taking $\Phi(f) = f^2$. If we set $g_{ij}(x,\dot{x}) = ff_{\dot{x}_i\dot{x}_j} + f_{\dot{x}_i}f_{\dot{x}_j}$, $Gi = -\frac{1}{2}Fi$, etc., the relation (3.4) is seen to have obvious connections with certain equations given by E. Cartan, "Les espaces de Finsler," *Actualités scientifiques et industrielles*, 79, *Exposés de Géométrie*, II, pp. 16 and 17.

and

$$P^*[t, x, \dot{x}] = f_{\dot{x}_i\dot{x}_j} X^i{}_{x_h} X^j{}_{x_k} \dot{x}_h \dot{x}_k + 2 f_{x_i\dot{x}_j} X^i X^j{}_{x_k} \dot{x}_k + f_{x_i x_j} X^i X^j \\ + f_{\dot{x}_i} \{ (X^i{}_{x_k} X^k)_{x_j} + X^i{}_{x_j t} \} \dot{x}_j + f_{x_i} \{ X^i{}_{x_j} X^j + X^i{}_t \}, \tag{3.6}$$

so that the following theorems may be stated in reasonably compact form.

THEOREM 4. *With the same assumptions as those of Theorem* 1 *regarding* C_0; C_1, $D(t)$, g_t, *we assume in addition that*

$$\phi(t) \leqq P[t, x, \lambda] \leqq \psi(t) \tag{3.7}$$

at every point $[x]$ *on* g_t *and for every set* $\lambda^1, \cdots, \lambda^n$ *such that* $f[x, \lambda] = 1$, *then*

$$\phi(t) D(t) \leqq D''(t) \leqq \psi(t) D(t). \tag{3.8}$$

THEOREM 5. *With the same assumptions as those of Theorem* 2 *regarding* C_0, C_1, Γ_t, $L(t)$, *we assume in addition that*

$$\phi^*(t) \leqq P^*[t, x, \lambda] \leqq \psi^*(t) \tag{3.9}$$

at every point $[x]$ *on* Γ_t *and for every set* $\lambda^1, \cdots, \lambda^n$ *such that* $f[x, \lambda] = 1$, *then*

$$\phi^*(t) L(t) \leqq L''(t) \leqq \psi^*(t) L(t). \tag{3.10}$$

In the Riemannian case we may write,

$$\begin{cases} B[\lambda] = [f[x, \lambda]]^3 P[t, x, \lambda] = Y_{ijkl}(x, t) \lambda^i \lambda^j \lambda^k \lambda^l \\ B^*[\lambda] = [f[x, \lambda]]^3 P^*[t, x, \lambda] = Y^*{}_{ijkl}(x, t) \lambda^i \lambda^j \lambda^k \lambda^l. \end{cases} \tag{3.11}$$

These biquadratic forms, in their relation to the appraisal of the second derivatives of $D(t)$ and $L(t)$ are analogous to the quadratic form $Q[\lambda]$ given by (2.4) in its relation to the appraisal of the first derivatives. There seems, however, to be no particularly simple interpretation for the coefficients Y_{ijkl} or $Y^*{}_{ijkl}$ such as we had for the X_{ij} in (2.4) as the covariant derivative of X_i.

To prove Theorem 4, differentiate (2.7) with respect to t and remember that $x_i = x_i(t, \tau)$, $i = 1, \cdots, n$, for $\tau = 0$ and 1 satisfy (1.1). We thus find that

$$D''(t) = K(t, 1) - K(t, 0) = \int_0^1 K_\tau(t, \tau)\, d\tau, \tag{3.12}$$

where

$$K(t,\tau)=f_{\dot{x}_i\dot{x}_j}[x(t,\tau),x_\tau(t,\tau)]X^j{}_{x_k}[x(t,\tau),t]\,X^i[x(t,\tau),t]\partial x_k/\partial\tau$$
$$+f_{\dot{x}_i x_j}X^iX^j+f_{\dot{x}_i}X^i{}_{x_j}X^j+f_{\dot{x}_i}X^i{}_t.$$

Now differentiating this last equality with respect to τ and using both (1. 2) and (3. 3) as well as the easily proved relations,

$$(\partial/\partial\tau)f_{\dot{x}_i\dot{x}_j}=f_{x_i\dot{x}_j}-f_{\dot{x}_i\dot{x}_k}F^k{}_{\dot{x}_j}-f_{\dot{x}_i x_j}=-\tfrac{1}{2}[f_{\dot{x}_i\dot{x}_k}F^k{}_{\dot{x}_j}+f_{\dot{x}_j\dot{x}_k}F^k{}_{\dot{x}_i}],$$
$$(\partial/\partial\tau)f_{\dot{x}_i x_j}=f_{x_i x_j}-f_{\dot{x}_i\dot{x}_k}F^k{}_{x_j},$$

we obtain

$$K_\tau(t,\tau)=P[t,x(t,\tau),x_\tau(t,\tau)].$$

Now, since $P[t,x,\dot{x}]$ is homogeneous of degree $+1$, we may introduce the parameter s as in the proof of Theorem 1, thus obtaining

$$K_\tau(t,\tau)=P[t,x(t,s),x_s(t,s)]f[x(t,\tau),x_\tau(t,s)],\qquad f[x(t,s),x_s(t,s)]=1.$$

From this, we find with the aid of (3. 8) and (2. 12), that

$$\phi(t)f[x,x_\tau]\leqq K_\tau(t,\tau)\leqq\psi(t)f[x,x_\tau].$$

Integrating from $\tau=0$ to $\tau=1$ and referring to (2. 6) and (3. 12), we find that (3. 8) has been proved, as desired.

To prove Theorem 5, we differentiate (2. 19) under the integral sign and make use of (1. 1). We thus obtain

$$L''(t)=\int_0^1 P^*[t,x(t,\tau),x_\tau(t,\tau)]d\tau.$$

Since $P^*[t,x,\dot{x}]$ is homogeneous of degree 1 in the $\dot{x}$'s, we have, upon introducing the arc length as a new parameter, the following formula:

$$L''(t)=\int_0^{L(t)} P^*[t,x(t,s),x_s(t,s)]ds.$$

It now follows immediately from (3. 9) that (3. 10) is true, thus completing the proof.

It is interesting to notice the difference between the two forms P and P^*:

$$P-P^*=f_{\dot{x}_i\dot{x}_j}\{X^iX^j{}_{x_k}F^k-X^iX^kF^j{}_{x_k}-\tfrac{1}{2}(X^iX^k)_{x_l}F^j{}_{\dot{x}_k}\dot{x}_l+X^iX^j{}_{x_hx_k}\dot{x}_h\dot{x}_k\}$$
$$+f_{x_i\dot{x}_j}\{X^i{}_{x_k}X^j-X^iX^j{}_{x_k}\}\dot{x}_k.$$

Upon rearranging the terms and writing

$$\begin{aligned} S_i = {} & f_{\dot{x}_i\dot{x}_j}\{X^j{}_{x_k}F^k - X^kF^j{}_{x_k} - \tfrac{1}{2}X^k{}_{x_l}F^j{}_{\dot{x}_k}\dot{x}_l + X^j{}_{x_h x_k}\dot{x}_h\dot{x}_k\} \\ & - \tfrac{1}{2}f_{\dot{x}_j\dot{x}_k}X^k{}_{x_l}F^j{}_{\dot{x}_i}\dot{x}_l + (f_{x_j\dot{x}_i}X^j{}_{x_k} - f_{x_i\dot{x}_j}X^j{}_{x_k})\dot{x}_k, \end{aligned} \tag{3.13}$$

we thus obtain

$$P - P^* = X^iS_i. \tag{3.14}$$

4. Special case of a Riemann metric with constant coefficients. For most purposes it is sufficient to take a Riemann metric with constant coefficients. If this is done, practically all formulas are enormously simplified. In this section we therefore assume that $f^2 = g_{ij}\dot{x}_i\dot{x}_j$, where the g's are constants. From (2.4), we have

$$Q[\lambda] = g_{ik}X^k{}_{x_j}\lambda^i\lambda^j. \tag{4.1}$$

Since f is independent of the undotted x's, we have $f_{x_i} \equiv f_{x_i\dot{x}_j} \equiv f_{x_ix_j} \equiv 0$, so that (3.4) yields $F^i \equiv 0$. From (3.5), we now obtain

$$\begin{aligned} f^3P = B[\dot{x}] = {} & (g_{\alpha\beta}g_{ij} - g_{\alpha i}g_{\beta j})(X^iX^j{}_{x_k})_{x_h}\dot{x}_\alpha\dot{x}_\beta\dot{x}_h\dot{x}_k \\ & + g_{\alpha\beta}g_{ih}(X^i{}_{x_j}X^j)_{x_k}\dot{x}_\alpha\dot{x}_\beta\dot{x}_h\dot{x}_k \\ & + g_{\alpha\beta}g_{ih}X^i{}_{tx_k}\dot{x}_\alpha\dot{x}_\beta\dot{x}_h\dot{x}_k. \end{aligned} \tag{4.2}$$

From (3.13), we have

$$f^3S_i = (g_{\alpha\beta}g_{ij} - g_{\alpha i}g_{\beta j})X^j{}_{x_kx_h}\dot{x}_\alpha\dot{x}_\beta\dot{x}_h\dot{x}_k. \tag{4.3}$$

Still further specializing to the Euclidean case we obtain

$$Q[\lambda] = X^i{}_{x_j}\lambda^i\lambda^j \tag{4.1E}$$

$$\begin{aligned} f^3P = {} & (X^iX^i{}_{x_k})_{x_h}\dot{x}_\alpha\dot{x}_\alpha\dot{x}_k\dot{x}_h - (X^iX^j{}_{x_k})_{x_h}\dot{x}_i\dot{x}_j\dot{x}_k\dot{x}_h \\ & + (X^i{}_{x_j}X^j)_{x_k}\dot{x}_\alpha\dot{x}_\alpha\dot{x}_i\dot{x}_k + X^i{}_{tx_k}\dot{x}_\alpha\dot{x}_\alpha\dot{x}_i\dot{x}_k. \end{aligned} \tag{4.2E}$$

$$f^3S_i = X^i{}_{x_hx_k}\dot{x}_\alpha\dot{x}_\alpha\dot{x}_h\dot{x}_k - X^j{}_{x_hx_k}\dot{x}_i\dot{x}_j\dot{x}_h\dot{x}_k. \tag{4.3E}$$

5. Second order linear systems. The simplest nontrivial system to which we may apply the above theory is the second order linear system,

$$dx/dt = A(t)x + B(t)y, \qquad dy/dt = C(t)x + D(t)y, \tag{5.1}$$

which we shall consider in connection with the metric,

$$ds^2 = Edx^2 + 2Fdxdy + Gdy^2, \qquad EG - F^2 > 0, \qquad E > 0. \tag{5.2}$$

Here E, F, G are constants. Thus, taking $g_{11} = E$, $g_{12} = F$, $g_{22} = G$, $\lambda^1 = \lambda$, $\lambda^2 = \mu$, $x_1 = x$, $x_2 = y$, the fundamental quadratic form Q may be computed from (4.1).

$$(5.3) \qquad Q = (AE + CF)\lambda^2 + (BE + DF + AF + CG)\lambda\mu + (BF + DG)\mu^2.$$

The extreme values of Q under the condition, $E\lambda^2 + 2F\lambda\mu + G\mu^2 = 1$, that λ and μ be the components of a unit vector, can be found by the method of Lagrange's multipliers. The maximum, $\beta(t)$, and the minimum, $\alpha(t)$, of $Q[\lambda]$ are thus seen to be the roots of the following quadratic equation in s:

$$(5.4) \qquad s^2 - (A + D)s + AD - BC - \tfrac{1}{4}(BE - AF + DF - CG)^2/(EG - F^2) = 0.$$

It is interesting to notice that the sum of the roots of this quadratic is the same as the sum of the roots of the equation,

$$(5.5) \qquad s^2 - (A + D)s + AD - BC = 0,$$

which, in the case of constant coefficients, A, B, C, D, is commonly called the characteristic equation. We shall call it the "characteristic equation," even when A, B, C, D are not constants. A *most favorable metric* (for some fixed value of t in the non-constant case) is defined as a metric which makes $\beta - \alpha$ a minimum. Since $\alpha + \beta = A + D$ is independent of the metric, it is clear that a most favorable choice also makes β a minimum and α a maximum. We summarize our results on most favorable metrics in the following three theorems:

THEOREM 6. *If*

$$(5.6) \qquad (A + D)^2 - 4(AD - BC) = (A - D)^2 + 4BC > 0,$$

a most favorable metric gives to α and β the values of the roots of the characteristic equation (5.5). *There exist infinitely many essentially distinct*[4] *most favorable metrics*

Proof. Since $(\beta - \alpha)^2$ is to be a minimum, we must choose E, F, and G so as to minimize the discriminant of (5.4). But the discriminant of (5.4)

---

[4] Two metrics, $Edx^2 + 2Fdxdy + Gdy^2$ and $E'dx^2 + 2F'dxdy + G'dy^2$, are for the purposes of Theorems 6 and 7 regarded as essentially the same, if, and only if,

$$E/E' = F/F' = G/G'.$$

is the same as the discriminant of (5.5) augmented by the non-negative quantity,

$$(BE-AF+DF-CG)^2/(EG-F^2).$$

Hence the required minimization is obtained by choosing E, F, and G, so that

$$(5.7)\qquad \begin{cases} BE-AF+DF-CG=0 \\ EG-F^2>0, \qquad E>0. \end{cases}$$

If $BC>0$, we satisfy (5.7) by taking $E=|C|$, $F=0$, $G=|B|$. If $BC<0$, we take $E=\pm(A-D)C$, $F=\pm 2BC$, $G=\pm(D-A)B$, where $\pm$ is chosen so that $\pm(A-D)C>0$. This makes $EG-F^2 = -BC[(A-D)^2+4BC]$, which in view of (5.6) must be positive in present circumstances. If $B=0$ and $C\neq 0$, take $E=1$, $F=\pm C$, $G=\pm(D-A)>0$. Similarly, if $B\neq 0$ and $C=0$. Finally, if $B=C=0$, take $E=G=1$ and $F=0$. From considerations of continuity, it is clear that in each of these cases there are infinitely many other choices.

THEOREM 7. *If*

$$(5.8)\qquad (A+D)^2-4(AD-BC)=(A-D)^2+4BC<0,$$

there exists an essentially unique most favorable metric, making $\alpha=\beta=$ *real part of one of the conjugate complex roots of the characteristic equation.*

Proof. In view of the known relationship between the roots of (5.5) and (5.4), it is sufficient to show that E, F, and G can be chosen in essentially just one way in such wise that (5.4) have a double root. Equating to zero the discriminant of (5.4) and making some algebraic manipulations, we find that E, F, and G must satisfy the relation,

$$(5.9)\qquad (BE+CG)^2-(2BF+DG-AG)(2CF+AE-DE)=0,$$

if α and β are to be equal. While there are obviously an infinite number of ways of satisfying this relation, it turns out that there is (neglecting a common factor) only one *real* set of values for E, F, G, namely

$$(5.10)\qquad E=\pm C>0,\quad F=\pm\tfrac{1}{2}(D-A),\qquad G=\mp B,$$

which satisfies (5.9).

In fact, if we set $p=2BF+DG-AG$ and $q=2CF+AE-DE$, we find at once that

$$(5.11)\qquad (BE+CG)(D-A)=Cp-Bq.$$

It follows from (5.9) that p and q must satisfy the relation

$$(5.12) \qquad (Cp - Bq)^2 - (D - A)^2 pq = 0.$$

The discriminant of this quadratic form in p and q is found to be $(A - D)^2[4BC + (A - D)^2]$, which, because of (5.8), is surely negative in case $A \neq D$. Since the left hand member of (5.12) is therefore positive definite, the only real values for p and q which can satisfy (5.12) are $p = q = 0$. The stated result follows at once, in case $A \neq D$. In the contrary case, $A = D$, we are justified, in view of (5.11) and the fact deduced from (5.8) that $BC \neq 0$, in writing $p = kB$ and $q = kC$. Thus (5.9) becomes $(BE + CG)^2 = k^2 BC$. The left hand side of this relation is non-negative, while, in view of (5.8), the right hand side is non-positive. Hence $k = 0$, and we are led to $p = q = 0$, as before, as well as to $BE + CG = 0$, from which the proof is readily completed.

THEOREM 8. *If*

$$(5.13) \qquad (A + D)^2 - 4(AD - BC) = (A - D)^2 + 4BC = 0,$$

there is no most favorable metric, except when $B = C = 0$. *But, for any number* $\epsilon > 0$, *a metric can always be found such that* $\frac{1}{2}(A + D) - \epsilon < \alpha \leqq \frac{1}{2}(A + D) \leqq \beta < \frac{1}{2}(A + D) + \epsilon$.

Proof. We prove the second statement first. If $A \neq D$, then $BC < 0$. Then, if we take $E = \pm (A - D)C > 0$, $F = \pm 2BC\theta$, $G = \pm (D - A)B$, $0 < \theta < 1$, we find that

$$\begin{aligned} EG - F^2 &= -BC(A - D)^2 - 4B^2C^2\theta^2 \\ &= -BC[(A - D)^2 + 4BC] - BC[4BC(\theta^2 - 1)], \end{aligned}$$

which, on account of (5.13) reduces to $4B^2C^2(1 - \theta^2) > 0$, so that $EG - F^2 > 0$, as required. We also find that equation (5.4) becomes

$$s^2 - (A + D)s + \tfrac{1}{4}(A + D)^2 - \tfrac{1}{4}[(A - D)^2(1 - \theta)]/(1 + \theta) = 0.$$

Since $0 < \theta < 1$, the roots of this equation are real; and, by taking θ sufficiently close to 1, we can make the roots differ from $\frac{1}{2}(A + D)$ by less than ϵ. If, however, $A = D$, either B or C is zero or both are zero. We now take $E = |C| + \eta$, $F = 0$, $G = |B| + \eta$, $\eta > 0$. It follows that $EG - F^2 > 0$, while (5.4) becomes

$$s^2 - 2As + A^2 = \tfrac{1}{4}[(B - C)^2\eta]/(|B| + |C| + \eta),$$

which yields real roots differing from A by less than ϵ, if η is sufficiently small.

To prove the first statement, we note that the established second statement implies that for any most favorable metric, $\alpha=\beta=\frac{1}{2}(A+D)$, a double root of (5.5). Thus, equations (5.4) and (5.5) must be identical equations, and thus we are led to the relation, $BE-AF+DF-CG=0$, to be satisfied by the coefficients of any most favorable metric. This relation can also be written in the form

$$(A-D)^2F^2=(BE-CG)^2. \tag{5.14}$$

From (5.13), we have (on multiplication by EG)

$$-(A-D)^2EG=4BCEG. \tag{5.15}$$

Adding corresponding members of (5.14) and (5.15), we obtain

$$-(A-D)^2(EG-F^2)=(BE+CG)^2\geqq 0.$$

Since $EG-F^2$ is to be positive, this relation is impossible unless $A=D$ and $BE=-CG$. From these two relations taken with (5.13), we find at once that we would have to have $B=C=0$ as stated.

We shall not dwell on the obvious connections between Theorems 6, 7, and 8 and the explicit solutions of (5.1) which are available when A, B, C, and D are constants. We merely remark that such considerations show that Theorems 1 and 2 are the best possible theorems of their type.

We now consider the biquadratic forms $B[\lambda]$ and $B^*[\lambda]=B-f^3X^iS_i$, whose significance is explained in **3**. We see at once from (4.3) that, for any linear system (not merely second order system) and for any Riemann metric with constant coefficients, all the S_i must vanish. Hence, our first result is to the effect that $B[\lambda]=B^*[\lambda]$.

In order to write down the expression for $B[\dot x]$ in reasonably compact and yet illuminating form, we introduce A^*, B^*, C^*, D^*, and H, defined as follows:

$$\begin{pmatrix}A^*B^*\\C^*D^*\end{pmatrix}=\begin{pmatrix}EF\\FG\end{pmatrix}\left[\begin{pmatrix}AB\\CD\end{pmatrix}^2+\begin{pmatrix}A_tB_t\\C_tD_t\end{pmatrix}\right]$$

$$H=EG-F^2.$$

We then find from (4.2) that

$$\begin{aligned}B[\dot x]&=H[-C\dot x^2+(A-D)\dot x\dot y+B\dot y^2]^2\\&\quad+[E\dot x^2+2F\dot x\dot y+G\dot y^2]\,[A^*\dot x^2+(B^*+C^*)\dot x\dot y+D^*\dot y^2].\end{aligned} \tag{5.16}$$

Hence, lower and upper bounds for B, for sets $(\dot{x}, \dot{y})$ satisfying $f^2 = E\dot{x}^2 + F\dot{x}\dot{y} + G\dot{y}^2 = 1$, can be written in the form $H\sigma^2 + \rho$ or ρ, where ρ being an extreme value of the quadratic form, $A^*\dot{x}^2 + (B^* + C^*)\dot{x}\dot{y} + D^*\dot{y}^2$, satisfies the quadratic equation,

$$H\rho^2 - [A^*G + D^*E - B^*F - C^*F]\rho + (A^*D^* - B^*C^*) - \tfrac{1}{4}(B^* - C^*)^2 = 0,$$

while σ, being an extreme value of the form, $-C\dot{x}^2 + (A - D)\dot{x}\dot{y} + B\dot{y}^2$, satisfies the quadratic

$$H\sigma^2 - [BE - AF + DF - CG]\sigma + (AD - BC) - \tfrac{1}{4}(A + D)^2 = 0.$$

In view of Theorem 7 and (5.10), the occurrence of the form, $\pm C\dot{x}^2 \mp (A - D)\dot{x}\dot{y} \mp B\dot{y}^2$, in (5.16) is rather interesting. It seems worth while therefore to investigate what happens when, for a certain value of t, (5.10) may be assumed to hold. The result depends upon the following matric identity, which the reader will have no trouble in verifying:

$$\operatorname{symm}\left[\begin{pmatrix} C & \tfrac{1}{2}(D - A) \\ \tfrac{1}{2}(D - A) & -B \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}^2\right] = \tfrac{1}{2}(A^2 + D^2 + 2BC)\begin{pmatrix} C & \tfrac{1}{2}(D - A) \\ \tfrac{1}{2}(D - A) & -B \end{pmatrix}.$$

Here the symbol symm M is used for the symmetric part of the matrix i. e. $\operatorname{symm} M = \tfrac{1}{2}(M + M')$. Using the definition of A^*, B^*, C^*, D^*, and H, we find from (5.16) that, when

$$E\dot{x}^2 + 2F\dot{x}\dot{y} + G\dot{y}^2 = \pm C\dot{x}^2 \pm (D - A)\dot{x}\dot{y} \mp B\dot{y}^2 = 1,$$

we necessarily have

$$\begin{aligned} B = \tfrac{1}{4}(A + D)^2 \pm \{[CA_t + \tfrac{1}{2}(D - A)C_t]\dot{x}^2 \\ + [CB_t - BC_t + \tfrac{1}{2}(D - A)(D_t + A_t)]\dot{x}\dot{y} \\ + [\tfrac{1}{2}(D - A)B_t - BD_t]\dot{y}^2\} = P[t, x, \dot{x}]. \end{aligned}$$

This result is fully to be expected in case the coefficients A, B, C and D of (5.1) are constants. Thus, if (5.10) holds for a particular value of t, it holds for all t. Our result shows that the estimates given in Theorems 4 and 5 are the best of their types. For we have exhibited an example in which the estimates give the exact values.

6. Application to a simple non-linear system. Consider the second order system,

$$dx/dt = -k(x^2+y^2)^{\frac{1}{2}}x$$
$$dy/dt = -g - k(x^2+y^2)^{\frac{1}{2}}y,$$

where k and g are positive constants, with the Euclidean metric, $ds^2 = dx^2 + dy^2$. Thus taking $g_{11} = g_{22} = 1$, $g_{12} = 0$, $x_1 = x$, $x_2 = y$, $\lambda^1 = \lambda$, $\lambda^2 = \mu$, the fundamental quadratic form Q may be computed from (4.1E). We thus find

$$-(1/k)Q[\lambda] = [x^2/(x^2+y^2)^{\frac{1}{2}} + (x^2+y^2)^{\frac{1}{2}}]\lambda^2 + (2xy/(x^2+y^2)^{\frac{1}{2}})\lambda\mu$$
$$+ [y^2/(x^2+y^2)^{\frac{1}{2}} + (x^2+y^2)^{\frac{1}{2}}]\mu^2.$$

If we set $v = (x^2+y^2)^{\frac{1}{2}}$, $\cos\theta = x/v$, $\sin\theta = y/v$, we obtain

$$\begin{aligned}-(1/kv)Q[\lambda] &= (1+\cos^2\theta)\lambda^2 + (2\cos\theta\sin\theta)\lambda\mu + (1+\sin^2\theta)\mu^2\\ &= \lambda^2+\mu^2+(\lambda\cos\theta+\mu\sin\theta)^2\\ &= 1+(\lambda\cos\theta+\mu\sin\theta)^2 \text{ for unit vectors } (\lambda,\mu),\end{aligned}$$

for which $\lambda^2 + \mu^2 = 1$. Evidently, for such unit vectors, Schwarz's inequality yields

$$0 \leqq (\lambda\cos\theta + \mu\sin\theta)^2 \leqq 1.$$

Therefore

$$1 \leqq -1/(kv)Q[\lambda] \leqq 2,$$
$$-2kv \leqq Q[\lambda] \leqq -kv.$$

We thus have an example of a system in which Q is negative definite at all points other than the origin. The biquadratic form B, however, seems too complicated to bother with.

7. Applications to the qualitative integration of differential equations. We close our paper with a few indications of the way in which our ideas may prove to be of value in the study of functions defined by systems of ordinary differential equations.

THEOREM 9. *Suppose that the region R of Theorem 1 is such that the $X^i[x,t]$ are defined for $[x]$ in R and for $t_0 \leqq t < +\infty$. Suppose also that* *no solution curve which for $t = t_0$ is within R ever passes through a boundary point of R for $t > t_0$. Let Λ be the least upper bound of*

$$f_{x_i}[x, \lambda]X^i[x, t] + f_{\dot{x}_i}[x, \lambda]X^i_{x_j}[x, t]\lambda^j \tag{7.1}$$

for $[x]$ in R for every set of λ's such that $f[x, \lambda] = 1$ and for $t \geqq t_0$.

Then, if $\Lambda < 0$, any two solutions, which for $t \geqq t_1 \geqq t_0$ are within R, must approach each other asymptotically. If R is finite, this phenomenon is uniform.

Proof. In fact, from either Theorem 1 or Theorem 2, it is obvious that the distance between corresponding points on the two trajectories is less than or equal to the distance at $t = t_1$, multiplied by $\exp[\Lambda(t - t_1)]$; $(\Lambda < 0)$. Hence it tends to zero as $t \to \infty$. The uniformity results from the fact that the distance at $t = t_1$ for any two trajectories is bounded, if R is finite. Further details are omitted.

THEOREM 10. *Suppose that, in addition to the hypothesis of Theorem 9, it is assumed that the X's depend periodically on t with period τ and that R is finite. Then there exists in R a unique periodic solution of period τ, toward which all other solutions entering R must approach asymptotically.*

Proof. Let $x_i = x_i^0(t)$, $i = 1, \cdots, n$, be an arbitrary solution which for $t \geqq t_0$ is within R. Then $x_i = x_i^m(t) = x_i^0(t + m\tau)$, $i = 1, 2, \cdots, n$, for each positive integral value of m, is also a solution. By the previous theorem, it is known that any two of these solutions approach each other uniformly. Hence, given a number $\epsilon > 0$ one can always find a number T, independent of p, such that

$$| x_i^0(t) - x_i^p(t) | < \epsilon \text{ for } t > T, \qquad p = 1, 2, \cdots.$$

In particular $| x_i^0(m\tau) - x_i^p(m\tau) | < \epsilon$ for integral $m > T/\tau$. From the definition of $x_i^m(t)$ this inequality can be written

$$| x_i^m(0) - x_i^{m+p}(0) | < \epsilon \text{ for } m > T/\tau.$$

Hence the points $[x^m(0)]$ form a Cauchy sequence and we can write

$$\lim_{m \to \infty} x_i^m(0) = a_i, \qquad i = 1, \cdots, n. \tag{7.2}$$

The solution, $x_i = x_i(t)$, such that $x_i(0) = a_i$ then turns out to be periodic. To establish this it is sufficient (on account of the uniqueness theorem for solutions of (1.1) taken together with the periodicity of the X's) to show that $x_i(\tau) = a_i$. By the known continuity of solutions in respect to their dependence on the initial conditions, we know that (7.2) implies

$$x_i(t) = \lim_{m\to\infty} x_i^m(t), \qquad i = 1, 2, \cdots, n,$$

uniformly over the finite interval $0 \leqq t \leqq \tau$. In particular,

$$x_i(\tau) = \lim_{m\to\infty} x_i^m(\tau) = \lim_{m\to\infty} x_i^{m+1}(0) = \lim_{m\to\infty} x_i^m(0) = a_i,$$

as we wished to prove. Finally, the periodic solution $[x(t)]$ must be unique. For, if there were two periodic solutions $[x(t)]$ and $[\bar{x}(t)]$ it is clear, from the fact that $\sum_{i=1}^{n} | x_i(t) - \bar{x}_i(t) |^2$ is continuous on $0 \leqq t \leqq \tau$, that this quantity has (on $0 \leqq t \leqq \tau$) a minimum value, which also serves (on account of periodicity) as the minimum value for all t. From Theorem 9, this minimum value must be 0. By the uniqueness theorem for differential equations, $x_i(t)$ and $\bar{x}_i(t)$ would then have to coincide, and the theorem is established.

The reader will notice that in the special case in which the X's do not depend explicitly on t, the periodic solution, mentioned in Theorem 10, must be an equilibrium point (i. e. a point where all the X's vanish), toward which all other solutions tend asymptotically. In fact, we can prove the following theorem for the non-existence of periodic solutions, which is somewhat reminiscent of a result of Bendixson in the case $n = 2$.[5]

THEOREM 11. *If the X's do not depend explicitly on t it is impossible to have a periodic solution (other than an equilibrium point) in any point set in which the expression* (7.1) *is either positive definite or negative definite.*

Proof. We consider only the case when (7.1) is negative definite. The other case reduces to this, if t is replaced by $-t'$. Suppose we had a given periodic solution $[x(t)]$ with period τ on which (7.1) were negative definite. Then, referring to Theorem 2, we take C_0 as the curve $x_i = x_i(t)$, C_1 as the curve $x_i = x_i(t + \tau)$, and Γ_{t_0} as the curve with one end point at $[x(t_0)]$ drawn around C_0 to the point $[x(t_0 + \tau)]$. Since $x_i(t_0 + \tau) = x_i(t_0)$, we see that Γ_{t_0}, and hence Γ_t, is closed. The length $L(t)$ is nothing other than the length of the given closed trajectory, which of course is a constant. Since

---

[5] Ivar Bendixson, "Sur les courbes définies par des équations différentielles," *Acta Mathematica*, vol. 24 (1901), pp. 1-88, especially p. 78.

(7.1) is negative definite, the β^* of (2.17) may be taken less than a fixed negative number. Here we make use of the fact that the closed trajectory is compact, but we omit details. Hence, from Theorem 2; $L(t) \to 0$ as $t \to \infty$. Hence $L(t) \equiv 0$. Thus the given periodic solution would reduce to an equilibrium point, and the theorem is proved.

In conclusion, we point out certain obvious connections between the results of this paper and a paper by Wintner.[6] We refer, for example, to his theorem to the effect that, if $A(t)$ is a matrix of n^2 real continuous functions satisfying

$$\text{(7.3)} \qquad \limsup_{m\to\infty} \int_0^t \max_{|y|=1} \, (yA(s)y)\,ds < \infty$$

then every solution vector $x = x(t)$ of $dx/dt = A(t)x$ is bounded as $t \to \infty$.

Here the quadratic form yAy plays the same role as our Q, and, in fact, this theorem of Wintner results immediately, if we apply our Theorem 1 or 2 to a linear system and a Euclidean metric.

UNIVERSITY OF MARYLAND, COLLEGE PARK, MD.
NAVAL ORDNANCE LABORATORY, WHITE OAK, MD.

---

[6] Aurel Wintner, "Asymptotic integration constants," *American Journal of Mathematics,* vol. 68 (1946), pp. 553-559, especially p. 558.

THE ASYMPTOTICALLY PERIODIC BEHAVIOR OF THE SOLUTIONS OF SOME LINEAR FUNCTIONAL EQUATIONS.*

By N. G. DE BRUIJN.

1. **Introduction.** A typical example of the equations to be dealt with in this paper is

$$x^{-\alpha}f'(x)+f(x)-f(x-1)=0, \tag{1.1}$$

where α is a positive constant. Any continuous[1] function $f(x)$, arbitrarily given in the interval $0\leqq x\leqq 1$, can be continued to a solution of (1.1) for $x\geqq 1$ simply by solving the differential equation $x^{-\alpha}g'(x)+g(x)=f(x-1)$ for $1\leqq x\leqq 2$ with the condition $g(1)=f(1)$, and so on.

It is easily seen that a function satisfying (1.1) for $x\geqq 1$ which is continuous for $x\geqq 0$ and bounded by M on $0\leqq x\leqq 1$ is also bounded by M on $0\leqq x<\infty$. Namely, let U be the least upper bound of $f(x)$ on $0\leqq x\leqq X$ and let x_0 be the smallest number $\geqq 0$ with $f(x_0)=U$. If x_0 would exceed 1, and so $f'(x_0)\geqq 0$, we could infer from (1.1) that $f(x_0-1)\geqq U$ and hence $f(x_0-1)=U$. This contradicts the minimum property of x_0; consequently $0\leqq x_0\leqq 1$ and $U=f(x_0)\leqq M$.

We can prove more, e. g. that $f(x)-f(x-1)\to 0$ for any solution of (1.1). If $\alpha>1$ we even find that $f(x)$ tends to a periodic function $\psi(x)$ of period 1; this function has derivatives of all orders and we have[2]

$$f^{(k)}(x)-\psi^{(k)}(x)\to 0 \text{ for } x\to\infty, \qquad k=0,1,2,\cdots. \tag{1.2}$$

If $0<\alpha\leqq 1$ this is no longer true. If $\frac{1}{2}<\alpha\leqq 1$ we still have a function $\psi(x)$ of period 1 such that

$$f^{(k)}(x)-\psi^{(k)}\{x-\int_1^x y^{-\alpha}dy\}\to 0 \text{ for } x\to\infty, \qquad k=0,1,2,\cdots. \tag{1.3}$$

If $0<\alpha\leqq\frac{1}{2}$ the asymptotic behavior of the solutions is not known. If, however, $\alpha\leqq 0$ it is comparatively simple: any solution tends to a constant. We shall deal with this last case, and with generalizations of it, in a separate paper.

---

* Received September 15, 1948.

[1] Throughout the paper all variables and functions are supposed to be real without explicit statement.

[2] If $f(x)$ is continuous for $x\geqq 0$ and satisfies (1.1) for $x\geqq 1$, then $f'(x)$ exists for $x\geqq 1$, f'' exists for $x\geqq 2$, etc.; cf. Lemma 2.

In the case of the equation (1.1) for $\alpha > \frac{1}{2}$, or more generally the homogeneous equations of class $\mathfrak{B}_n$ ($n \geqq 2$) below, we apparently have a linear operator $Tf = \psi$; its domain[3] consists of the functions $f(x)$ which are continuous on $0 \leqq x \leqq 1$ and its range $\boldsymbol{R}$ consists of a set of periodic functions of period 1. This operator is continuous: if $f_n \to f$ uniformly on $0 \leqq x \leqq 1$, then we have uniformly $Tf_n \to Tf$. We shall prove furthermore that[4] $\boldsymbol{R}$ is *dense* in the space $\boldsymbol{R}_0$ of continuous periodic functions of period 1, that is to say, if a function $\chi \varepsilon \boldsymbol{R}_0$ and a positive number ϵ are arbitrarily given, then a solution $f(x)$ of the functional equation can be found such that $\psi = Tf$ satisfies $|\psi(x) - \chi(x)| < \epsilon$ for all values of x.

The following uniqueness problem arises: Is the correspondence $f \to \psi$ one to one? Or, what is the same thing, does $f \to 0$ for $x \to \infty$ imply $f \equiv 0$? The answer is affirmative for a rather special class of equations (**5**), including the equation (1.1) for $\alpha > 1$, but the result probably holds true for wider classes. Thus far we know no counter examples, as far as equations (2.1) of class $\mathfrak{B}_n$ ($n \geqq 2$) with $q(x) > 0$, $r(x) \equiv 0$ are concerned. For the equation of **6**, Example 3, which actually gave rise to the present investigation, the uniqueness theorem is also true, although the conditions of **5** are not satisfied. We do not prove it in this paper; it requires entirely different methods.

Throughout the paper, in conditions and statements of the type "$f^{(k)}(x)$ is continuous for $x \geqq a$," $f^{(k)}(a)$ must be interpreted as the right-hand k-th derivative.

2. Functional equations of class $\mathfrak{A}_n$.

We want to study equations of the type $w(x)f'(x) + f(x) - f(x-1) = 0$, where $w(x)$ is a positive function. The methods used here compel us, however, to consider the more general non-homogeneous equation (2.1).

We introduce an arbitrary decreasing positive function $\Phi(x)$ for $x \geqq 1$ with convergent integral $\int_1^\infty \Phi(x)\,dx$. Let A be a positive number such that

$$A > \Psi(1), \qquad A > \int_1^\infty \Psi(x)\,dx.$$

---

[3] It is possible of course to extend our results to a wider domain, but such an extension would not lead to more essential generalizations. If, for instance $f(x)$ is absolutely integrable over $0 \leqq x \leqq 1$ and if (1.1) is interpreted reasonably then $f(x)$ is continuous for $1 \leqq x \leqq 2$. Our results are not essentially influenced by a shift $x \to x + 1$.

[4] The restriction $q(x) > 0$ has to be made here; see Theorem 6, Remark 1.

If B is a positive and D a non-negative number, the equation

$$(2.1) \qquad w(x)f'(x) + p(x)f(x) - q(x)f(x-1) = r(x)$$

is called an equation of class $\mathfrak{A}_0(\Phi, B, D)$ if the functions w, p, q, r are continuous for $x \geqq 1$ and if, furthermore

$$(2.2) \qquad w(x) > 0 \quad (x \geqq 1)$$

$$(2.3) \qquad p(x) > \tfrac{1}{2} \quad (x \geqq B+1)$$

$$(2.4) \qquad |p(x) - q(x)| < B\Phi(x), \qquad |r(x)| \leqq D\Phi(x) \qquad (x \geqq 1).$$

The equation (2.1) is said to belong to class $\mathfrak{A}_n(\Phi, B, D)$ if it furthermore satisfies the following additional conditions: the functions w, p, q, r have a continuous derivative for $x \geqq 1$, a continuous second derivative for $x \geqq 2, \cdots$, and a continuous n-th derivative for $x \geqq n+1$, which satisfy

$$(2.5) \quad |w^{(k)}| < B\Phi, \qquad |p^{(k)}| < B\Phi, \qquad |q^{(k)}| < B\Phi, \qquad |r^{(k)}| \leqq D\Phi$$

for $x \geqq k+1$, $k = 1, 2, \cdots, n$.

Evidently we have $\mathfrak{A}_0 \supset \mathfrak{A}_1 \supset \mathfrak{A}_2 \supset \cdots \supset \mathfrak{A}_n$. An equation is said to belong to the class $\mathfrak{A}_\infty(\Phi)$ if, for any $n \geqq 0$, numbers B_n and D_n can be found such that it belongs to $\mathfrak{A}_n(\Phi, B_n, D_n)$.

The class $\mathfrak{B}_n(\Phi, B, D)$ consists of all equations of class $\mathfrak{A}_n(\Phi, B, D)$ which satisfy

$$(2.6) \qquad \{w(x)\}^2 < B\Phi(x) \qquad (x \geqq 1).$$

Evidently (1.1) belongs to $\mathfrak{A}_\infty(x^{-\alpha-1})$ if $\alpha > 0$ and to $\mathfrak{B}_\infty(x^{-\beta})$, where $\beta = \mathrm{Min}(2\alpha, \alpha+1)$, if $\alpha > \frac{1}{2}$.

Henceforth symbols $C_1, C_2, \cdots$ will appear which have to be interpreted as follows. If such a C occurs in a formula or statement, then it can be given a positive value such that the formula or statement is valid; the choice of that value may depend on the function $w(x)$ and on the values of A and B only, certainly not on x, $f(x)$, $\Phi(x)$, $p(x)$, $q(x)$, $r(x)$ and D. If, however, a variable is written down explicitly, e. g. $C(u)$, then the choice may depend also on that variable. Analogously we shall use symbols $\gamma_1, \gamma_2, \cdots$ which depend on A and B only, and not on $w(x)$.

The proofs of the relations (1.2) or (1.3) do not require this convention about the C's, but more explicit estimates are needed in **4**; it requires only little extra trouble to derive them here.

A function $f(x)$ is said to be a *solution* of (2.1) if it is continuous for $x \geqq 0$, derivable for $x \geqq 1$ and if it satisfies (2.1) for $x \geqq 1$.

LEMMA 1. *If* (2.1) *belongs to* $\mathfrak{A}_0(\Phi, B, D)$ *and if the function* $g(x)$ *is continuous for* $0 \leqq x \leqq 1$, *then a uniquely determined solution* $f(x)$ *can be found with* $f(x) = g(x)$ $(0 \leqq x \leqq 1)$.

Proof. Since $w(x)$, $p(x)$, $r(x)$ and $g(x-1)$ are continuous for $1 \leqq x \leqq 2$, the differential equation

$$w(x)f'(x) + p(x)f(x) = q(x)g(x-1) + r(x)$$

has a solution in the interval $1 \leqq x \leqq 2$ with $f(1) = g(1)$. This process can be continued.

LEMMA 2. *If the integers* k *and* n *satisfy* $0 \leqq k \leqq n$, *and if* (2.1) *belongs to* $\mathfrak{A}_n(\Phi, B, D)$, *then any solution* $f(x)$ *has* $k+1$ *continuous derivatives for* $x \geqq k+1$.

Proof. We have

$$f'(x) = \{q(x)f(x-1) + r(x) - p(x)f(x)\}/w(x) \qquad (x \geqq 1). \tag{2.7}$$

If $f(x)$ has k $(k \leqq n)$ continuous derivatives for $x \geqq k$ then the right-hand side of (2.7) has the same property for $x \geqq k+1$. Consequently $f(x)$ has $k+1$ continuous derivatives for $x \geqq k+1$. Our lemma follows by induction.

LEMMA 3. *If the conditions of the previous lemma are satisfied, and if*

$$|f(x)| \leqq M \qquad (0 \leqq x \leqq 1), \tag{2.8}$$

then we have, for $k = 0, 1, \cdots, n+1$ *and any* $s > 0$,

$$|f^{(k)}(x)| \leqq (M + D) \cdot C_1(s, k) \qquad (k \leqq x \leqq s). \tag{2.9}$$

Proof. First take $k = 0$, and put

$$q(x)f(x-1) + r(x) = \phi(x) \qquad (x \geqq 1). \tag{2.10}$$

The function $y = f(x)$ solves the differential equation

$$w(x)y' + p(x)y = \phi(x), \tag{2.11}$$

and so we find, for $x \geqq 1$, $\xi \geqq 1$,

$$\begin{aligned} f(x) = \exp\{-\int_\xi^x p(t)/w(t)\,dt\} \cdot f(\xi) \\ + \int_\xi^x \phi(t)/w(t) \cdot \exp\{-\int_t^x p(u)/w(u)\,du\}\,dt. \end{aligned} \tag{2.12}$$

Take $\xi = 1$, $1 \leqq x \leqq 2$; since $|\phi(t)| \leqq (M+D)C_2$ for $1 \leqq t \leqq 2$, we obtain

$$|f(x)| \leqq (M+D)C_3 \qquad (1 \leqq x \leqq 2).$$

Now taking $\xi = 2$, $2 \leqq x \leqq 3$, we can apply the same process and we find an estimate for f on $2 \leqq x \leqq 3$. Thus (2.9) can be proved for $k = 0$.

After that, (2.7) gives the result for $k = 1$. By differentiation of (2.7) we derive the result for $k = 2$, etc.

THEOREM 1. *If* (2.1) *belongs to* $\mathfrak{A}_0(\Phi, B, D)$ *and if a solution* $f(x)$ *satisfies* (2.8), *then*

$$(2.13) \qquad |f(x)| \leqq (M+D)C_4 \qquad (0 \leqq x < \infty).$$

Proof. Let B_1 $(B_1 > 0)$ be such that $p(x) > \frac{1}{2}$ for $x > B_1 + 1$; $B_1 = B$ suffices. Put

$$M^*(x) = \operatorname*{Max}_{0 \leqq t \leqq x} f(t); \qquad M(x) = \operatorname*{Max}_{0 \leqq t \leqq x} |f(t)|.$$

If $x \geqq 1$ we either have $M^*(x) = M^*(x-1)$ or there is a number x_0 $(x - 1 \leqq x_0 \leqq x)$ such that $M^*(x) = f(x_0)$, $f'(x_0) \geqq 0$. In the latter case we have, by (2.1)

$$p(x_0)f(x_0) \leqq q(x_0)f(x_0 - 1) + r(x).$$

It results from (2.2) that, for $x \geqq B_1 + 2$

$$M^*(x) = f(x_0) \leqq M^*(x-1) + 2\,|M^*(x-1)|\,B\Phi(x-1) + 2D\Phi(x-1).$$

This is, of course, also true in the first case. For $\operatorname{Min} f(t)$ an analogous inequality can be found; it follows that

$$M(x) \leqq M(x-1) \cdot \{1 + 2B\Phi(x-1)\} + 2D\Phi(x-1),$$

and so

$$D + M(x) \leqq (D + M(x-1)) \cdot \{1 + 2(B+1)\Phi(x-1)\} \qquad (x \geqq B_1 + 2).$$

Since $\Phi(1) < A$, $\int_1^\infty \Phi(x)\,dx < A$, we have $(0 \leqq B_1 \leqq B)$

$$\prod_{\nu=1}^{\infty} \{1 + 2(B+1)\Phi(B_1 + \nu)\} < \gamma_1 \qquad (\gamma_1 > 1),$$

and hence

$$(2.14) \qquad |f(x)| \leqq \{M(B_1 + 1) + D\}\gamma_1 \qquad (x \geqq B_1).$$

We may take $B_1 = B$. By Lemma 3 we obtain

$$|f(x)| \leqq (M+D)C_1(B+1,0) = (M+D)C_5 \quad (0 \leqq x \leqq B+1)$$

and (2.13) directly follows, with $C_4 = \gamma_1 \cdot (C_5 + 1)$.

If $B_1 = 0$ we do not need Lemma 3 for the proof of Theorem 1. Then (2.14) expresses

THEOREM 1a. *If the conditions of Theorem 1 are satisfied and if* $p(x) > \frac{1}{2}$ *for* $x \geqq 1$, *then a number* γ_1, *depending on* A *and* B *only, exists such that*

$$|f(x)| \leqq (M+D)\gamma_1 \qquad (x \geqq 0).$$

LEMMA 4. *If the conditions of Lemma 3 are satisfied and if* $n \geqq 1$, *then for* $1 \leqq k \leqq n$ *the function* $f_k(x) = f^{(k)}(x+k)$ *is a solution of an equation*

$$w_k(x)f'_k(x) + p_k(x)f_k(x) - q_k(x)f_k(x-1) = r_k(x) \tag{2.15}$$

which belongs to $\mathfrak{A}_{n-k}(\Phi_k, B_k, D_k)$, *where*

$\Phi_k(x) = \Phi(x+k)$, $w_k(x) = 2^k w(x+k)$, $B_k = C_6(n)$, $D_k = (M+D)C_7(n)$.

Proof. Let N be a natural number and suppose that the lemma has been proved for $n < N$.[5] Then we have, by Lemma 3 and Theorem 1

$$|f^{(k)}(x)| \leqq (M+D)C_8(N) \qquad (x \geqq k,\ k = 0, 1, \cdots, N-1). \tag{2.16}$$

Now suppose that $n = N$.

It is evidently sufficient to consider the case $k = 1$; the cases $k = 2, 3, \cdots$ immediately follow from this one by iteration.

Since $f(x)$ satisfies (2.1) we obtain by differentiation, for $x \geqq 2$,

$$\begin{aligned} w(x)f''(x) + \{p(x) + w'(x)\}f'(x) - q(x)f'(x-1) \\ = r'(x) - p'(x)f(x) + q'(x)f(x-1). \end{aligned}$$

Now putting

$$\left\{ \begin{array}{ll} 2w(x+1) = w_1(x), & 2\{p(x+1) + w'(x+1)\} = p_1(x) \\ 2q(x+1) = q_1(x), & \\ \quad 2\{r'(x+1) - p'(x+1)f(x+1) + q'(x+1)f(x)\} = r_1(x), & \end{array} \right. \tag{2.17}$$

the function $f_1(x) = f'(x-1)$ satisfies (2.15), with $k = 1$, for $x \geqq 1$. We have to prove that the analogues of (2.3), (2.4) and (2.5) hold.

First we have, since $|w'(x)| < B\Phi(x) \to 0$ for $x \to \infty$:

$$p_1(x) > \tfrac{1}{2} \text{ for } x \geqq C_9 + 1.$$

---

[5] If $N = 1$ nothing is supposed here.

We immediately find

$$|w_1^{(l)}(x)| < 2B\Phi_1(x), \quad |q_1^{(l)}(x)| < 2B\Phi_1(x) \quad (l = 1,\cdots,N-1, x \geqq l+1),$$

and hence, for $x \geqq l+1$,

$$|\{p_1(x) - q_1(x)\}^{(l)}| < 2B\Phi_1(x) + 2\,|w^{(l+1)}(x+1)| < 4B\Phi_1(x) \qquad (l = 0, 1, \cdots, N-1)$$

$$|p_1^{(l)}(x)| < 6B\Phi_1(x) \qquad (l = 1, \cdots, N-1).$$

Finally, by (2.17), (2.16), (2.4), and (2.5) we have, for $l = 0, 1, \cdots, N-1$, $x \geqq l+1$,

$$|r_1^{(l)}(x)| < 2\{D + 2B(M+D)\cdot 2^l C_8(N)\}\cdot\Phi(x).$$

Now by taking

$$C_6(N) = \text{Max}\,(C_9, 6B), \qquad C_7(N) = 2 + 2^{N+1}BC_8(N)$$

the proof of the case $n = N$, $k = 1$ is completed.

We state separately the result (2.16):

THEOREM 2. *If* $0 \leqq k \leqq n$, *and if* $f(x)$ *is a solution of an equation* $\mathfrak{A}_n(\Phi, B, D)$ *with* $|f(x)| \leqq M$ $(0 \leqq x \leqq 1)$, *then we have*

$$|f^{(k)}(x)| \leqq (M+D)C_{10}(k) \qquad (x \geqq k, k = 0, 1, \cdots, n).$$

We may write $C_{10}(k)$ instead of the $C_8(n)$ from (2.16) since $\mathfrak{A}_n(\Phi, B, D)$ is contained in $\mathfrak{A}_k(\Phi, B, D)$.

In **4** we shall need

LEMMA 5. *If* $f(x)$ *satisfies the equations* (2.1) *of class* $\mathfrak{A}_2(\Phi, B, D)$ *and if a number* $Q \geqq 2$ *is given such that*

$$p(x) > \tfrac{1}{2}, \qquad p(x) + 2w'(x) > \tfrac{1}{2} \qquad (x \geqq Q+1)$$

$$|f^{(k)}(x)| \leqq M \qquad (k = 0, 1, 2;\ Q \leqq x \leqq Q+1),$$

then a number γ_2, *depending on* A *and* B *only, can be found such that*

$$|f^{(k)}(x)| \leqq (M+D)\gamma_2 \qquad (k = 0, 1, 2;\ x \geqq Q).$$

Proof. For $k = 0$ this follows by applying Theorem 1a to $f(x-Q)$. This same theorem can be applied to $f'(x-Q)$ which satisfies an equation of class $\mathfrak{A}_1(\Phi_Q, B_1, D_1)$, where $B_1 = \gamma_3$, $D_1 = (M+D)\gamma_4$ (cf. the proof of Lemma 4). A third application yields the result for $k = 2$.

THEOREM 3. *Under the conditions of Theorem* 2 *we have, if* $n \geqq 1$, $x \geqq k$, $x \geqq B+1$,

$$(2.18) \qquad |f^{(k)}(x) - f^{(k)}(x-1)| \leqq C_{11}(k)\cdot\{w(x) + \Phi(x)\}(M+D).$$

Proof. We have by (2.1)

$$(2.19)\quad p(x)\{f(x)-f(x-1)\} = \{q(x)-p(x)\}f(x-1)+r(x)-w(x)f'(x).$$

Hence for $x \geqq B+1$ we have by (2.2), etc. and Theorem 2

$$2\,|\,f(x)-f(x-1)\,| \leqq C_{12}(n)\cdot\{w(x)+\Phi(x)\}(M+D).$$

We may write C_{13} instead of $C_{12}(n)$; cf. the remark after Theorem 2. This proves (2.18) for $k=1$. The cases $k=2,3,\cdots$ are easily dealt with by successive differentiation of (2.19).

3. Periodicity Theorem. If $w(x)\to 0$ for $x\to\infty$, Theorem 3 yields $f(x)-f(x-1)\to 0$. It by no means implies that $f(x)$ tends to a periodic function. But this is easily proved if we suppose that

$$(3.1)\qquad |\,w(x)\,| < B\Phi(x) \qquad (x \geqq 1)$$

and that $f(x)$ satisfies an equation (2.1) of class $\mathfrak{A}_1(\Phi, B, D)$. We then infer from (2.18) that

$$|\,f(x+\nu)-f(x+\mu)\,| \leqq (M+D)C_{14}\int_{x+\mu}^{\infty}\Phi(t)\,dt$$
$$(\nu,\mu = 1,2,3,\cdots;\ \mu\leqq\nu,\ x\geqq 0).$$

Hence, by Cauchy's theorem, $\lim\limits_{\nu\to\infty} f(x+\nu)$ exists.

We, however, drop this restriction (3.1) and replace it by the weaker one

$$(3.2)\qquad \{w(x)\}^2 < B\Phi(x), \qquad (x \geqq 1);$$

that is, our equation belongs to a class $\mathfrak{B}_n$.

LEMMA 6. *If $f(x)$ satisfies an equation (2.1) of class $\mathfrak{B}_2(\Phi, B, D)$ and if (2.8) holds, then there exists a uniquely determined continuous periodic function $\psi(y)$ of period 1 such that for $x \geqq C_{21}$*

$$(3.3)\qquad |\,f(x)-\psi\{x-\int_{B+1}^{x} w(t)/p(t)\,dt\}\,| \leqq (M+D)C_{15}\int_{x-1}^{\infty}\Phi(t)\,dt.$$

Proof. By Theorem 2, $f''(x)$ is uniformly bounded for $x \geqq 2$. So it follows from (3.2) that, for $x \geqq B+2$

$$|\,f\{x+w(x)/p(x)\}-f(x)-w(x)/p(x)\cdot f'(x)\,| \leqq (M+D)C_{16}\Phi(x),$$

where $C_{16}=2BC_{10}(2)$. Consequently, by (2.1)

$$(3.4)\quad \begin{cases} |\,f\{x+w(x)/p(x)\}-f(x-1)\,| \leqq \{p(x)\}^{-1}\cdot|\{q(x)-p(x)\}f(x-1)+r(\\ \qquad +\,(M+D)C_{16}\Phi(x) \leqq (M+D)C_{17}\Phi(x) \qquad (x \geqq B+2) \end{cases}$$

For $x > C_{18}$ the function

$$y = x - \int_{B+1}^{x} w(t)/p(t)\,dt$$

is steadily increasing, and $dy/dx \to 1$. We put $y = \eta(x)$, $x = \xi(y)$. Take $x_0 > C_{18} + 1$ and put $y_0 = 1 + \eta(x_0 - 1)$, $x^* = \xi(y_0)$, so that

$$x^* - x_0 = \int_{x_0-1}^{x^*} w(t)/p(t)\,dt. \tag{3.5}$$

Since

$$|\{w(x)/p(x)\}'| < C_{19}\Phi(x) \qquad (x \geqq B + 1),$$

we easily infer from (3.5) that

$$|x^* - x_0 - w(x_0)/p(x_0)| < C_{20}\Phi(x_0 - 1) \qquad (x_0 \geqq C_{21} + 1). \tag{3.6}$$

The function $f'(x)$ is also uniformly bounded, and so by (3.6) and (3.4)

$$\begin{aligned} |f\{\xi(y_0)\} - f\{\xi(y_0 - 1)\}| &= |f(x^*) - f(x_0 - 1)| \\ &\leqq |f\{x_0 + w(x_0)/p(x_0)\} - f(x_0 - 1)| + (M + D)C_{10}(1) \cdot C_{20}\Phi(x_0 - 1) \\ &\leqq (M + D)\,C_{22}\Phi(x_0 - 1). \end{aligned}$$

Since $\int_1^\infty \Phi(x)\,dx < \infty$ we find that

$$\psi(y) = \lim_{\nu\to\infty} f\{\xi(y + \nu)\}$$

exists if ν runs through the natural numbers, and (3.3) immediately follows. Evidently $\psi(y)$ has the period 1. Moreover, from (3.3) and from the fact that $f'(x)$ is bounded, it is easily deduced that

$$|\psi(y_1) - \psi(y_2)| \leqq (M + D)\,C_{23} \cdot |y_1 - y_2| \tag{3.7}$$

for any pair of values y_1, y_2; hence $\psi(y)$ is continuous.

THEOREM 4. *If $f(x)$ is a solution of the equation* (2.1) *of class* $\mathfrak{B}_{n+2}(\Phi, B, D)$, $(n \geqq 0)$ *then there exists a uniquely determined periodic function $\psi(y)$, with period* 1 *and with n successive derivatives, such that for* $k = 0, 1, \cdots, n$; $x \geqq k + C_{24}(k)$

$$|f^{(k)}(x) - \psi^{(k)}(x - \int_{B+1}^{x} w(t)/p(t)\,dt)| \leqq (M + D)\,C_{25}(k) \int_{x-1}^{\infty} \Phi(t)\,dt. \tag{3.8}$$

Proof. Lemma 4 shows that $f^{(k)}(x + k)$ satisfies an equation of class $\mathfrak{B}_2$, so that, by Lemma 5, continuous functions $\psi_0, \psi_1, \cdots, \psi_n$ exist such that for $x \geqq k + C_{24}(k)$,

$$|f^{(k)}(x) - \psi_k(x - \int_{\lambda}^{x} w(t)/p_k(t)\,dt)| \leqq (M + D)\,C_{26}(k) \int_{x-1}^{\infty} \Phi(t)\,dt. \tag{3.9}$$

Here we have $p_k(t) = p(t) + kw'(t)$ (cf. the proof of Lemma 4), and λ is chosen such that $p(t), p_1(t), \cdots, p_n(t)$ are all $> \frac{1}{3}$ for $x \geqq \lambda$. Since for $x \geqq \lambda$

$$| J_k(x) | = | \int_x^\infty \{w(t)/p(t) - w(t)/p_k(t)\} dt < | 9k\{B\Phi(1)\}^{\frac{1}{2}} \cdot B \int_x^\infty \Phi(t)\, dt,$$

(cf. (3. 2)) we have, in virtue of (3. 7), for $k = 0, 1, \cdots, n$,

$$(3.10) \qquad | f^{(k)}(x) - \phi_k(x - \int_{B+1}^x w/p\, dt) | \leqq (M + D) C_{27}(k) \int_{x-1}^\infty \Phi(t)\, dt,$$

where

$$\phi_k(u) = \psi_k\{u + J_k(\lambda) + \int_{B+1}^\lambda w/p\, dt\}.$$

By integration of (3. 10) we obtain, using the facts that the ϕ_k's are continuous and that $w(x)/p(x) \to 0$ for $x \to \infty$,

$$\int_\alpha^\beta \phi_k(x)\, dx = \phi_{k-1}(\beta) - \phi_{k-1}(\alpha) \qquad (k = 1, 2, \cdots, n)$$

for any α and β. It results that

$$\phi'_0 = \phi_1,\ \phi'_1 = \phi_2, \cdots,\ \phi'_{n-1} = \phi_n.$$

4. The correspondence $f \to \psi$; the completeness theorem. We consider an equation (2. 1) of the type $\mathfrak{B}_n$ $(n \geqq 2)$. Any continuous function $f(x)$ arbitrarily given for $0 \leqq x < 1$ gives rise to a continuous function $\psi(x)$ of period 1 in virtue of Lemma 6. We shall investigate this correspondence. Since it is linear we may restrict ourselves to the homogeneous case $r(x) \equiv 0$. Then the correspondence $f \to \psi$ defines a linear operator $\psi = Tf$. The domain is the set $\boldsymbol{D}$ of all continuous functions on $0 \leqq x \leqq 1$, or, since any such function can be continued to a solution, the set of all solutions of our equation. The range $\boldsymbol{R}$ is the set of all $\psi = Tf$, $f \varepsilon \boldsymbol{D}$. By $\boldsymbol{R}_0$ we denote the set of all continuous functions of period 1. Using the metric:

$$(4.1) \qquad \| f \| = \operatorname*{Max}_{0 \leqq x \leqq 1} | f(x) |, \qquad \| \psi \| = \operatorname*{Max}_{0 \leqq x \leqq 1} | \psi(x) |,$$

we have

THEOREM 5. *The operator T is continuous in $\boldsymbol{D}$*:

$$(4.2) \qquad \| Tf \| \leqq C_4 \| f \| \qquad (f \varepsilon \boldsymbol{D}).$$

Proof. This immediately follows from (2. 13) (where $D = 0$ since our equation is homogeneous) and from Lemma 6.

THEOREM 6. (*Completeness Theorem*). *If* $q(x) > 0$ *for* $x \geqq 1$, *then* $\boldsymbol{R}$ *is dense in* $\boldsymbol{R}_0$, *that is to say, if* $\chi \varepsilon \boldsymbol{R}$ *and* $\epsilon > 0$ *are given, then a function* $f \varepsilon \boldsymbol{D}$ *can be found such that* $\| \chi - Tf \| < \epsilon$.

Remarks. 1. The condition $q(x) > 0$ for $x \geqq 1$ is essential. If we had, for instance, $q(x) = 0$ for $1 \leqq x \leqq 2$, then any pair of solutions f_1 and f_2 with $f_1(1) = f_2(1)$ would coincide for $x \geqq 1$, so that $\boldsymbol{R}$ would consist of all multiples of a single function.

2. Theorem 6 is also true if an equation of class $\mathfrak{A}_1$ is concerned, with the supplementary condition (3.1); only very little alterations in the proof below will be required.

LEMMA 7. *Let the equation* (2.1) *be of class* $\mathfrak{A}_0$, *with* $q(x) > 0$ $(x \geqq 1)$. *For* $Q \geqq 0$ *denote the set of continuous functions on* $Q \leqq x \leqq Q + 1$ *by* $\boldsymbol{D}_Q$. *Let for* $0 \leqq R \leqq Q$ *the operator* T_{RQ} *be defined as follows: if* $g \varepsilon \boldsymbol{D}_Q, h \varepsilon \boldsymbol{D}_Q$, *and if the function* $f(x)$ *satisfies* (2.1) *for* $x \geqq R + 1$, f *continuous for* $x \geqq R$, $f \equiv g$ $(R \leqq x \leqq R + 1)$, $f \equiv h$ $(Q \leqq x \leqq Q + 1)$ *then we put* $h = T_{RQ}g$. *Then* $T_{RQ}\boldsymbol{D}_R$ *is dense*[6] *in* $\boldsymbol{D}_Q$.

Proof. It results from (2.12), with $\xi = R + 1$, $r(x) \equiv 0$, that the operator T_{RQ} is continuous. Furthermore, for $P \leqq R \leqq Q$, we have $T_{RQ}T_{PR} = T_{PQ}$. Thus, if $T_{PR}\boldsymbol{D}_P$ is dense in $\boldsymbol{D}_R$ and $T_{RQ}\boldsymbol{D}_R$ is dense in $\boldsymbol{D}_Q$, then $T_{PQ}\boldsymbol{D}_P$ is dense in $\boldsymbol{D}_Q$. Consequently it suffices to prove our lemma for the case $R \leqq Q \leqq R + 1$; the general case follows by iteration of this one.

Let $R \leqq Q \leqq R + 1$, $\epsilon > 0$ and $h \varepsilon \boldsymbol{D}_Q$. Construct a function $h_1 \varepsilon \boldsymbol{D}_Q$ with a continuous first derivative, such that

$$w(x)h'_1(x) + p(x)h_1(x) - q(x)h_1(x-1) = 0 \tag{4.3}$$

holds for $x = Q + 1$, and $\| h_1 - h \| < \epsilon$. Since $q(x) > 0$ we can continue $h_1(x)$ backwards over $R \leqq x \leqq Q$ such that it satisfies (4.3) for $R + 1 \leqq x \leqq Q + 1$ and such that it is continuous for $R \leqq x \leqq Q + 1$.

If $g \varepsilon D_R$ is defined by $g(x) = h_1(x)$ $(R \leqq x \leqq R + 1)$, then we have $T_{RQ}g = h_1$, and so $\| T_{RQ}g - h \| < \epsilon$, which proves the lemma.

Proof of Theorem 6. Let M_0 be the largest of the numbers $\| \chi \|$, $\| \chi' \|$ and $\| \chi'' \|$. Let ϵ_1 $(0 < \epsilon_1 < 1)$ be a number which is to be fixed later on. Let the number $Q > B + 3$ be such that for $x \geqq Q - 1$ we have

$$\begin{cases} p(x) > \frac{1}{4}, & p(x) + 2w'(x) > \frac{1}{4}, \qquad q(x) > \frac{1}{4} \\ w(x) < \epsilon_1, & \Phi(x-1) < \epsilon_1, \qquad \int_x^\infty \Phi(t)\,dt < \epsilon_1, \end{cases} \tag{4.4}$$

---

[6] With the metric defined analogous to (4.1).

and such that

$$N = Q - \int_{B+1}^{Q} w(t)/p(t)\,dt \tag{4.5}$$

is an integer. It follows from Lemma 7, with $R = Q - 2$, that a continuous function $f_1(x)$ can be found in $Q - 2 \leqq x \leqq Q + 1$ which solves (2.1) for $Q - 1 \leqq x \leqq Q + 1$, such that

$$| f_1(x) - \chi(x - Q) | < \epsilon_1 \qquad (Q \leqq x \leqq Q + 1). \tag{4.6}$$

Again by Lemma 7, with $R = 0$, a solution $f(x)$ of (2.1) can be found such that

$$| f(x) - f_1(x) | < \epsilon_1 \cdot \operatorname*{Max}_{Q-1 \leqq x \leqq Q+1} p(x)\{w(x)\}^{-2} \qquad (Q - 2 \leqq x \leqq Q - 1). \tag{4.7}$$

Putting $f(x) - f_1(x) = f_2(x)$ we have

$$f'_2(x) = \{q(x) f_2(x-1) - p(x) f_2(x)\}/w(x) \qquad (Q - 1 \leqq x \leqq Q + 1),$$

and a similar expression for f''_2 $(x \geqq Q)$. From the inequalities (2.4), (2.5), (4.4), and from $p(x) > \frac{1}{4}$ we now easily infer that

$$| f^{(k)}(x) - f_1^{(k)}(x) | < \gamma_5 \epsilon_1, \qquad (k = 0, 1, 2;\ Q - 2 + k \leqq x \leqq Q - 1 + k) \tag{4.8}$$

γ_5 depending on A and B only. Consequently, by Lemma 5,

$$| f^{(k)}(x) | \leqq (M_0 + \gamma_5 \epsilon_1 + \epsilon_1)\gamma_2 = M_1 \qquad (k = 0, 1, 2;\ x \geqq Q). \tag{4.9}$$

We can now derive by the process of Lemma 6[7]

$$| f\{x + w(x)/p(x)\} - f(x - 1) | \leqq \gamma_6 M_1 \Phi(x) \qquad (x \geqq Q + 1)$$

and, if $Tf = \psi$,

$$| f(x) - \psi\{x - \int_{B+1}^{x} w/p\,dt\} | \leqq \gamma_7 M_1 \int_{x-1}^{\infty} \Phi(t)\,dt \qquad (x \geqq Q). \tag{4.10}$$

It follows from (4.9) and from Theorem 4 $(k = 1)$ that $\| \psi' \| \leqq M_1$. So (4.10) and (4.5) lead to

$$| f(x) - \psi(x - Q + N) | < \epsilon_1 M_1 + \gamma_7 M_1 \epsilon_1 \qquad (Q \leqq x \leqq Q + 1),$$

and from (4.6), (4.8) we infer

$$\| \chi - \psi \| < \epsilon_1\{\gamma_5 + 1 + (1 + \gamma_7)(M_0 + \gamma_5 + 1)\gamma_2\} = \epsilon_1 \gamma_8(\chi).$$

Now take ϵ_1 such the right-hand side equals ϵ; then it results that $\| \chi - \psi \| = \| \chi - Tf \| < \epsilon$.

---

[7] We can take $C_{18} = C_{21} = Q$, and then for $x \geqq Q$ all other C's can be replaced by γ's.

5. The uniqueness problem. We consider the equation

$$w(x)f'(x)+p(x)f(x)-f(x-1)=0. \tag{5.1}$$

Let B and ρ be positive constants, $\rho>1$, and suppose that for $x\geqq 1$ $(n=0,1,2,\cdots)$, the functions $w^{(n)}(x)$ and $p^{(n)}(x)$ are continuous and satisfy

$$|w^{(n)}(x)|<B^{n+1}n^nx^{-n-\rho},\quad |\{p(x)-1\}^{(n)}|<B^{n+1}n^nx^{-n-\rho}\quad (0^0=1). \tag{5.2}$$

Under these conditions we can prove

THEOREM 7. *If* $f(x)$ *is a solution of* (5.1) *and if* $\lim\limits_{x\to\infty}f(x)=0$, *then we have* $f(x)\equiv 0$.

Remarks. 1. The condition $w(x)>0$ $(x\geqq 1)$ is not needed here. If it is valid, however, then the equation (5.1) belongs to $\mathfrak{B}_\infty(x^{-\rho})$ and gives rise to a one-to-one correspondence $f\leftrightarrows\psi$.

2. With very little alterations in the proof below we can show that Theorem 6 remains true if the right-hand sides of (5.2) are replaced by

$$B^nn^{\alpha n}x^{-\beta n-\rho},$$

where the numbers B, α, β, ρ satisfy

$$B>0,\quad \rho>1,\quad \rho>\alpha,\quad \rho>\alpha-\beta+1.$$

For instance, the function $w(x)=\exp(-e^x)$ satisfies this condition, with $\alpha=3$, β and ρ arbitrary; it does not satisfy (5.2).[8]

LEMMA 8. *If* $a\geqq 0$, *if* $f(x)$ *is a solution of* (5.1), *and if* $g(x)$ *satisfies the "adjoint equation"*

$$\{w(x)g(x-1)\}'-p(x)g(x-1)+g(x)=0 \tag{5.3}$$

for $x\geqq a+1$, $g(x)$ *continuous for* $x\geqq a$, *then the expression*

$$(f,g)=\int_{x-1}^{x}f(t)g(t)\,dt+w(x)f(x)g(x-1) \tag{5.4}$$

is independent of x *for* $x\geqq a+1$.

This directly follows by differentiation of (5.4) with respect to x.

---

[8] Condition (5.2) implies that $w(x)$ and $p(x)$ can be continued analytically throughout a region $|\arg x|<\delta_1$, $|x|>1-\delta_2$ $(\delta_{1,2}>0)$ and that $w(x)=O(x^{-\rho})$, $p(x)-1=O(x^{-\rho})$, in that region. Conversely, from these facts we can deduce that (5.2) is valid for some $B>0$.

LEMMA 9. (*Theorem of Carleman-Mandelbrojt*[9]). *If* $a < b$ *and* $\eta > 1$, *then a function* $g_0(x)$ *exists for* $a \leqq x \leqq b$, *differentiable infinitely often, such that for* $n = 0, 1, 2, \cdots$

$$(5.5)\quad \begin{cases} |\,g_0^{(n)}(x)\,| < n^{\eta n} \quad (a \leqq x \leqq b\,;\, 0^0 = 1), \quad g_0^{(n)}(a) = g_0^{(n)}(b) = 0, \\ g_0(x) \geqq 0 \quad (a \leqq x \leqq b), \quad \int_a^b g_0(x)\,dx > 0. \end{cases}$$

Proof of Theorem 7. Suppose that $f(x)$ is not identically zero, then numbers a and b $(0 \leqq a < b \leqq a + 1)$ can be found such that $f(x) \neq 0$ $(a \leqq x \leqq b)$. Take $1 < \eta < \rho$. Let $g_0(x)$ be the function of Lemma 9, put $g_0(x) = 0$ for $b \leqq x \leqq a + 1$, and let $g_n(x)$ be defined recursively by

$$(5.6)\qquad g_n(x) = (1 - \Omega_n)\, g_{n-1}(x),$$

where the operator Ω_n is defined by

$$(5.7)\qquad \Omega_n \phi = \alpha_n(x) \cdot \phi + \{w_n(x)\phi\}',$$

$$(5.8)\qquad \alpha_n(x) = 1 - p(x + n), \qquad w_n(x) = w(x + n).$$

Furthermore we define $g(x)$ for $x \geqq a$ by

$$(5.9)\qquad g(x + n) = g_n(x) \qquad (a \leqq x \leqq a + 1, n = 0, 1, 2, \cdots).$$

This $g(x)$ clearly satisfies (5.3) for $x \geqq a + 1$; we shall prove that $g(x)$ is bounded for $x \geqq a$.

First we estimate the expression $\Omega_{k_s}\Omega_{k_{s-1}} \cdots \Omega_{k_1} g_0$ for $1 \leqq k_1 < \cdots < k_s$, $a \leqq x \leqq a + 1$. We write it as a polynomial in the α's, the w's, g_0 and their derivatives, and we then observe, by induction, that all its terms are contained in the expansion of

$$(5.10)\qquad (1 + d/dx)^s \{(\alpha_{k_s} + w_{k_s}) \cdots (\alpha_{k_1} + w_{k_1}) g_0\}.$$

We can write this in the form

$$(1 + \mu + \mu_{k_1} + \cdots + \mu_{k_s})^s \{(\alpha_{k_s} + w_{k_s}) \cdots (\alpha_{k_1} + w_{k_1}) g_0\},$$

where μ denotes the differential operator working on g_0 only, μ_j the one working on α_j and w_j only. We now replace the α's, w's, g_0 and their derivatives by the majorants given by (5.2), (5.5) and we even replace m^m for $m < s$ by s^m. We infer that the sum of the absolute values of the terms in (5.10) is less than

---

[9] See S. Mandelbrojt, *Séries de Fourier et classes quasi-analytiques de fonctions*, Paris, 1935, p. 69; the above lemma is a direct application. Mandelbrojt's contribution, viz., the fact that $g_0(x)$ can be chosen $\geqq 0$, is essential in the sequel.

We do not actually need the Carleman-Mandelbrojt theorem here since it can be verified that $g_0(x) = \exp\{-(x - a)^{1/(\eta-1)} - (b - x)^{1/(\eta-1)}\}$ $(a < x < b)$, $g_0(a) = g_0(b) = 0$ satisfies the conditions of the above lemma.

$$U = B(k_1 \cdots k_s)^{-\rho} \cdot (1 + s^\eta + 2s/k_1 + \cdots + 2s/k_s)^s,$$

and a fortiori we obtain

$$(5.11)\begin{cases} |\Omega_{k_s} \cdots \Omega_{k_1} g_0| \leqq U \leqq B(k_1 \cdots k_s)^{-\rho}(A_1 s^\eta)^s \\ \quad \leqq B(k_1 \cdots k_s)^{-\rho} A_2{}^s \Gamma(1 + \eta s) = B(k_1 \cdots k_s)^{-\rho} \int_0^\infty e^{-t}(A_2 t^\eta)^s dt; \end{cases}$$

$A_1, A_2, \cdots$ denote positive constants depending on B, ρ and η only.

From (5.6) and (5.11) it results that

$$\begin{aligned} |g_n(x)| &= |(1 - \Omega_n) \cdots (1 - \Omega_1) g_0| \\ &\leqq B \int_0^\infty e^{-t} \prod_{k=1}^n (1 + A_2 k^{-\rho} t^\eta)\, dt \leqq B \int_0^\infty e^{-t} \prod_{k=1}^n (1 + A_3 k^{-\rho/\tau} t^{\eta/\tau})^\tau dt. \end{aligned}$$

where $\tau = \frac{1}{2}(\rho + \eta)$. Finally we obtain

$$|g_n(x)| \leqq B \int_0^\infty \exp(-t + A_3 \tau t^{\eta/\tau} \sum_1^\infty k^{-\rho/\tau})\, dt.$$

The integral on the right is convergent (since $\rho > \tau$, $\eta < \tau$) and its value is independent of n. Consequently $|g(x)| < A_4$ for $x \geqq a$.

The proof of Theorem 7 now easily follows from Lemma 8. From $f(x) \to 0$, $|g(x)| < A_4$ for $x \to \infty$ we obtain by (5.4) that $(f, g) = 0$. On the other hand, by taking $x = a + 1$ in (5.4), we find

$$(f, g) = \int_a^b f(t) g(t)\, dt.$$

But since $f(x)$ is continuous and $\neq 0$ for $a \leqq x \leqq b$, $g(x) \geqq 0$, $g(x)$ not identically 0, we find $(f, g) \neq 0$. This is contradictory; it follows that $f \equiv 0$.

6. Applications.

We shall apply the contents of this paper to some equations of the form

$$(6.1) \qquad G'(x) = K(x) G(x - 1),$$

where $K(x)$ is a given function, positive and continuous for $x \geqq 1$.

If $G_0(x)$ is a solution of (6.1) and $G_0(x) > 0$ for $0 \leqq x \leqq 1$, then clearly $G_0(x) > 0$ for $x \geqq 0$. Moreover, if $G(x)$ is a second solution, then $G(x)/G_0(x)$ is bounded for $x \geqq 0$. Namely, if M is the maximum of

$|G(x)/G_0(x)|$ for $0 \leqq x \leqq 1$, then it is easily seen by step-by-step integration that $|G(x)| \leqq MG_0(x)$ for $x \geqq 0$.

If we put $G(x) = G_0(x)f(x)$, the function $f(x)$ satisfies

$$f'(x)(G_0(x)/G'_0(x)) + f(x) - f(x-1) = 0, \tag{6.2}$$

which belongs to class $\mathfrak{A}_0$.

Example 1. The equation

$$G'(x) = 2xe^{2x-1}G(x-1). \tag{6.3}$$

A special solution is $G_0(x) = \exp x^2$, and so (6.2) becomes

$$(2x)^{-1}f'(x) + f(x) - f(x-1) = 0.$$

This equation belongs to class $\mathfrak{B}_\infty$ with $\Phi(x) = x^{-2}$. Hence any solution of (6.3) gives rise to a periodic function $\psi(x)$ of period (1), such that

$$G(x) = e^{x^2}\{\psi(x - \tfrac{1}{2}\log x) + O(x^{-1})\}.$$

The completeness theorem is valid, but the uniqueness problem is still unsolved.

Example 2. The equation

$$G'(x) = (\exp x^2)G(x-1). \tag{6.4}$$

This example is less artificial than the previous one; it is not easy to give an explicit solution $G_0(x)$. Nevertheless we can try to find a function $G_0^*(x) = e^{H(x)}$ such that the substitution $G(x) = G_0^*(x)f(x)$ leads to an equation of a class $\mathfrak{B}_n$ $(n \geqq 2)$. We find, if $H'(x) = h(x)$,

$$\begin{aligned} f'(x)/h(x) &+ f(x) \\ &- \exp\{x^2 - \int_0^1 h(x-t)\,dt - \log h(x)\}f(x-1) = 0, \end{aligned} \tag{6.5}$$

and hence we have to find a function $h(x)$ which makes the absolute value of

$$x^2 - \int_0^1 h(x-t)\,dt - \log h(x) \tag{6.6}$$

less than a function $\Phi(x)$ with convergent integral. The following iteration process gives a result in a finite number of steps in this and in many other cases. Put

$$h_0(x) = 1, \quad h_1(x) = x^2, \quad h_{n+1}(x) = h_n(x) + \Delta_{n+1}(x)$$

$$\Delta_{n+1}(x) = \Delta_n(x) - \int_0^1 \Delta_n(x-t)\,dt - \log\{h_n(x)/h_{n-1}(x)\},$$

so that $\Delta_{n+1}(x)$ is the result of substituting $h(x) = h_n(x)$ into (6.6). Since we shall find $w(x) = \{h(x)\}^{-1} \sim x^2$ we cannot apply our theory with a function Φ of order less than x^{-3} (cf. (2.5)), therefore we stop our iteration at the first Δ_n which is $O(x^{-3})$.

We obtain

$$\begin{aligned}
\Delta_2(x) &= x - \tfrac{1}{3} - 2\log x \\
\Delta_3(x) &= \tfrac{1}{2} - 2x^{-1} + 2x^{-2} + 2x^{-2}\log x - 2x^{-3}\log x + O(x^{-3}) \\
\Delta_4(x) &= \tfrac{1}{2}x^{-2} - 2x^{-3}\log x + O(x^{-3}) \\
\Delta_5(x) &= O(x^{-3})
\end{aligned}$$

$$h_4(x) = x^2 + x + \tfrac{1}{6} - 2x^{-1} + x^{-2} - 2\log x + 2x^{-2}\log x - 4x^{-3}\log x + O(x^{-3}).$$

Now take

$$\begin{aligned}
(6.7)\quad H(x) = \tfrac{1}{3}x^3 + \tfrac{1}{2}x^2 + \tfrac{13}{6}x - 3x^{-1} - 2x\log x - \\
- 2\log x - 2x^{-1}\log x + 2x^{-2}\log x,
\end{aligned}$$

so that $H'(x) = h(x) = h_4(x) + O(x^{-3})$. Now the equation (6.5) belongs to $\mathfrak{B}_\infty(x^{-3})$; the required inequalities (2.5) for the derivatives of $w(x)$ and $q(x)$ easily follow from the fact that these functions are power series in x^{-1} and $x^{-1}\log x$. It results that to any solution of (6.4) there belongs a periodic function $\psi(x)$, of period 1, having derivatives of all orders, with

$$G(x) = e^{H(x)}\{\psi(x + x^{-1}) + O(x^{-2})\},$$

or

$$G(x) = e^{H(x)}\{\psi(x) + O(x^{-1})\}.$$

$H(x)$ is given by (6.7). The completeness theorem holds; the uniqueness theorem can be proved by applying Theorem 7 to the equation (6.5); first divide it by the coefficient of $f(x-1)$.

Example 3. Mahler's partition problem[10] gives rise to the functional equation $F'(y) = F(y/r)$, where r is a constant > 1. We transform it by $y = r^x$, $F(r^x) = G(x)$ into

$$(6.8)\qquad G'(x) = e^{\alpha x + \beta}G(x-1) \qquad (\alpha > 0),$$

---

[10] K. Mahler, "On a special functional equation," *Journal of the London Mathematical Society*, vol. 15 (1940), pp. 115-123.

N. G. de Bruijn, "On Mahler's partition problem," *Proceedings of the Koninklijke Nederlandsche Akademie van Wetenschappen*, Amsterdam, vol. 51 (1948), pp. 659-669 = *Indagationes Mathematicae*, vol. 10 (1948), pp. 210-220.

where especially $\alpha = \log r$, $\beta = \log\log r$. Dealing with (6.8) by the method of the preceding example, we find

$$h_0(x) = 1, \qquad h_1(x) = \alpha x + \beta$$

$$\Delta_2(x) = \tfrac{1}{2}\alpha - \log\alpha - \log x - \beta\alpha^{-1}x^{-1} + O(x^{-2})$$

$$\begin{aligned}\Delta_3(x) = {} & x^{-1}(-1 + \alpha^{-1}\log\alpha) + \alpha^{-1}x^{-1}\log\alpha \\ & + \alpha^{-2}(-\beta - \tfrac{1}{2}\alpha + \log\alpha)x^{-2}\log x + \tfrac{1}{2}\alpha^{-2}x^{-2}\log^2 x + O(x^{-2})\end{aligned}$$

$$\Delta_4(x) = -\tfrac{1}{2}\alpha^{-1}x^{-2}\log x - \alpha^{-2}x^{-2}\log x + O(x^{-2})$$

$$\Delta_5(x) = O(x^{-2})$$

$$(6.9) \qquad \left\{\begin{aligned} H(x) = {} & \tfrac{1}{2}\alpha(x - \alpha^{-1}\log x)^2 + (1 + \beta + \tfrac{1}{2}\alpha - \log\alpha)x \\ & + (-1 + \alpha^{-1}\log\alpha - \beta/\alpha)\log x - \tfrac{1}{2}\alpha^{-2}x^{-1}\log^2 x \\ & + \alpha^{-2}(\alpha + \beta - \log\alpha)x^{-1}\log x.\end{aligned}\right.$$

To every solution of (6.8) there corresponds a function $\psi(y)$ of period 1, with derivatives of all orders, such that[11]

$$(6.10) \qquad G(x) = e^{H(x)}\{\psi(x - \log x/\alpha + \log x/\alpha^2 x) + O(x^{-1})\};$$

the rather complicated function $H(x)$ is given by (6.9). Section **4** applies to this case, section **5** does not. Nevertheless the uniqueness theorem is true; this will be shown in a separate paper dedicated specially to (6.8). There we shall express $\psi(x)$ explicitly in terms of the values of $G(x)$ on $0 \leqq x \leqq 1$, and vice versa.

From the results to be obtained in that paper it follows that (6.10) is the best possible result of its kind: it is not true that functions $H_1(x) = O\{H(x)\}$ and $\eta(x) \to \infty$ exist such that every solution of (6.8) is of the form

$$G(x) = e^{H_1(x)}(\psi\{\eta(x)\} + o(x^{-1})),$$

$\psi(y)$ periodic mod 1.

Mathematisch Instituut der Technische Hogeschool,
Delft, Netherlands.

---

[11] We take $\Phi(x) = x^{-2}$.

ON THE CLASSICAL EXISTENCE THEOREM OF LINEAR DIFFERENTIAL EQUATIONS.*

By Aurel Wintner.

The purpose of the following considerations is, first to formulate, and then to prove, the classical representation theorem of homogeneous, linear systems of ordinary differential equations so as to apply to absolutely convergent Dirichlet series or Laplace integrals.

Theorem. *On the half-line*

$$1 \leqq t < \infty, \tag{1}$$

let $\alpha(t)$, $\beta(t)$ *denote n by n matrices of functions each of which is of bounded variation on* (1), *and put*

$$A(s) = \int_1^\infty e^{-st} d\alpha(t). \tag{2}$$

and

$$B(s) = \int_1^\infty e^{-st} d\beta(t), \tag{3}$$

where

$$0 \leqq s < \infty. \tag{4}$$

There belongs to every α *a* β *having the following property: n linearly independent solution vectors* $x(s)$ *of the system of n linear differential equations*

$$x' = A(s)x, \tag{5}$$

where $x' = dx/ds$, *are supplied by the n columns of the matrix*

$$X(s) = E + B(s), \tag{6}$$

where

$$E \text{ is the unit matrix.} \tag{7}$$

It should be emphasized that the absolute convergence of the unknown

---

* Received September 3, 1948.

matrix integral, (3), is claimed on the entire half-line, (4), on which the absolute convergence of the given matrix integral, (2), is assumed.

It will be clear from the proof (cf. (27) below) that the assertion of the theorem can be amplified by the following

REMARK. *Under the assumptions of the Theorem, the spectrum of β is contained in any closed ray containing the spectrum of α.*

It is understood that the spectrum of α is defined to be the set of those points t_0 of the half-line (1) corresponding to which there does not exist an $\epsilon > 0$ satisfying $\alpha(t) = \text{const.}$ for $|t - t_0| < \epsilon$, and that a ray, R, of a set, S, is meant to be a set containing S and having the property the *sum* of any two (not necessarily distinct) values contained in R is in R.

In order to illustrate the content of the Theorem (and of the Remark), suppose that each of the n^2 elements of the coefficient matrix $A(s)$ of a system (5) is a Dirichlet series without constant term, say

$$d(s) = \sum_{m=1}^{\infty} a_m e^{-\lambda_m s}, \text{ where } 0 < \lambda_1 < \lambda_2 < \cdots; \tag{8}$$

that each of these Dirichlet series is absolutely convergent on (4); finally, that (by the insertion of vanishing coefficients a_m) the notation is so chosen that there belongs to every h and every j $(\leqq h)$ an m satisfying $\lambda_h + \lambda_j = \lambda_m$. It then follows that every component $x_i = x_i(s)$ of every solution vector $x = (x_1, \cdots, x_n)$ of (5) is of the form $x_i(s) = c_i + d_i(s)$, where $c_1 = x_1(\infty), \cdots, c_n = x_n(\infty)$ are integration constants, which can be assigned arbitrarily, and $d_1(s), \cdots, d_n(s)$ are Dirichlet series having the same exponents λ_m as (8) and converging on the entire half-line (4).

The particular case $\lambda_m = m$ is classical. For, if the n^2 elements of $A(s)$ belong to the case $\lambda_m = m$ of (8), the substitution $e^{-s} = z$ reduces (5) to

$$dw/dz = P(z)w, \tag{10}$$

where each of the n^2 elements of $P(z)$ is a regular power series,

$$\sum_{m=0}^{\infty} b_m z^m \tag{11}$$

(*where* $b_0 \neq 0$ *is allowed*). Hence, what results is the classical result, according to which the regularity of the coefficient matrix (11) in a circle $|z| < r$ implies the regularity of every component $w_i = w_i(z)$ of every solution $w = (w_1, \cdots, w_n)$ of (10) in the entire circle.[1]

---

[1] What actually follows is somewhat more, namely, the absolute convergence of the power series of every function $w_i(z)$ on the *boundary* $|z| = r$, if the power series (11),

The classical proofs of the latter theorem fail to apply in the case of arbitrary Dirichlet series, $\lambda_m \neq m$, mainly because the regularity of the functions ceases to imply the convergence (not to say the *absolute* convergence) of the Dirichlet expansions. In fact, this difficulty arises even in the case of ordinary Dirichlet series,

$$d(s) = \sum_{m=2}^{\infty} a_m/m^s, \tag{12}$$

where, in (8),

$$\lambda_m = \log (m+1). \tag{13}$$

It is of no avail that (12) cannot have a half-plane of convergence without having some half-plane of absolute convergence.

The proof of the Theorem proceeds as follows:

Under the assumptions placed on the coefficient matrix, $A(s)$, of (5) in the Theorem, put

$$[\alpha] = \max ([\alpha_{11}], [\alpha_{12}], \cdots, [\alpha_{nn}]), \tag{14}$$

where, if $\alpha_{ik}(t)$ denotes the (i, k)-th element of the matrix $\alpha(t)$,

$$[\alpha_{ik}] = \int_1^{\infty} | d\alpha_{ik}(t) | . \tag{15}$$

Then the assumption placed on (2) is

$$[\alpha] < \infty. \tag{16}$$

It can be assumed that

$$\alpha(1) = 0 \tag{17}$$

($0 =$ zero matrix), since (2) remains unaltered if a constant matrix is added to the matrix $\alpha(t)$.

Let X be an n-ary square matrix the columns of which are solution vectors of (5). Then (5) is equivalent to

$$X'(s) = A(s)X(s). \tag{18}$$

Hence, if $B(s)$ is defined by (6) and (7),

$$B'(s) = A(s) + A(s)B(s). \tag{19}$$

---

representing the coefficient functions of (11) when $|z| < r$, are supposed to converge absolutely on $|z| = r$.

This refinement of the classical theorem can readily be verified directly.

Since $B(s)$ is required to become of the form (3), it is required that

$$B(\infty) = 0. \tag{20}$$

Hence, the successive approximations $B_1(s), B_2(s), \cdots$, which should lead, in the form

$$B(s) = \sum_{m=1}^{\infty} B_m(s), \tag{21}$$

to the desired solution, (3), of (19), must be set up as follows:

$$B_{m+1}'(s) = A(s) B_m(s), \tag{22}$$

where $m = 1, 2, \cdots$ and

$$B_1'(s) = A(s). \tag{23}$$

In order to obtain every term of (21) in a form corresponding to (3), try

$$B_m(s) = \int_1^{\infty} e^{-st} d\beta_m(t), \tag{24}$$

where $m = 1, 2, \cdots$. The unknown matrices $\beta_m(t)$ occurring in (24) can be normalized by

$$\beta_m(1) = 0; \tag{25}$$

cf. (17).

According to the case $m = 1$ of (24), the initial requirement, (23), of the recursion formula, (22), is satisfied if β_1 is defined (at its continuity points) by (25) and

$$-t d\beta_1(t) = d\alpha(t). \tag{26}$$

For $\beta_2, \beta_3, \cdots$, substitution of (24) into (22) supplies the recursion formula

$$-t d\beta_{m+1}(t) = d\{\alpha(t) * \beta_m(t)\}, \tag{27}$$

where the asterisk is the symbol of convolutions (of matrices; so that $\lambda * \mu$ is *not* identical with $\mu * \lambda$). Needless to say, only $\beta_m(t+0)$ or $\beta_m(t-0)$ is relevant in (26), (27).

It is readily seen from (15), and from the definition of a convolution as an integral, that, if $\lambda_{ik}(t)$ and $\mu_{jk}(t)$ are scalar functions,

$$[\lambda_{ik} * \mu_{jk}] \leqq [\lambda_{ik}][\mu_{jk}].$$

Hence, if $\lambda(t)$ and $\mu(t)$ are n-rowed square matrices, it follows from (14) and (15) that

$$[\lambda * \mu] \leqq n[\lambda][\mu], \tag{28}$$

the factor n being introduced by the summations which occur in a matrix product. On the other hand, $t \geqq 1$, by (1), and so, by (26),

$$[\beta_1] \leqq [\alpha]. \tag{29}$$

For the same reason,

$$[\beta_{m+1}] \leqq [\alpha * \beta_m],$$

by (27). Consequently, an induction and (29) show that

$$[\beta_m] \leqq (n[\alpha])^m. \tag{30}$$

It follows from (29), (30) and (16) that (26) and (27) define on the half-line (1) matrix functions each of which has a finite total variation on (1). Thus, each of the integrals (24) is defined, and converges absolutely, on the half-line (4).

On the other hand, (29) and (30) fail to lead, via (24) and (21), to an integral representation (3) in which, as claimed by the Theorem, $[\beta] < \infty$. In fact, all that follows in this manner is the absolute convergence of (3) on *some* half-line $s_0 \leqq s < \infty$, with a sufficiently *large* s_0, rather than on the *given* half-line (4), where $s_0 = 0$. Correspondingly, what is thus supplied by (30) is of the type of the local existence theorem of *non-linear* analytic differential equations.

Thus, in order to prove the Theorem, the *linear* character of (5) will have to be used more fully than above. This will be accomplished by proving that (30) can be refined to

$$[\beta_m] \leqq (n[\alpha])^m/m!, \tag{31}$$

where $m = 1, 2, \cdots$.

To this end, use will be made of the notion of the spectrum, as defined after the italicized Remark following the Theorem. It is clear from this notion, and from the integral defining a convolution $\lambda * \mu$ of two matrices $\mu = \mu(t)$, $\lambda = \lambda(t)$ on (1), that a t-value, say t_0, cannot be in the spectrum of $\lambda * \mu$ unless t_0 is in the closure of the set of those numbers which are representable in the form $t_1 + t_2$, where t_1, t_2 are points contained in the spectra of λ, μ respectively. This implies that, if the spectra of λ, μ are contained in the respective half-lines $a \leqq t < \infty$, $b \leqq t < \infty$, where $a \geqq 0$,

$b \geqq 0$, then the spectrum of $\lambda * \mu$ is contained in the half-line $a + b \leqq t < \infty$.

Since all spectra are confined to the half-line (1), it now follows from (26) and (27), by choosing

$$\lambda = \alpha, \qquad \mu = \beta_m, \qquad\qquad a = 1, b = m$$

and applying an induction, that the spectrum of β_m is contained in the half-line $m \leqq t < \infty$, where $m = 1, 2, \cdots$. In view of (25) and of the definition of a spectrum, this means that

$$\beta_m(t) \equiv 0 \text{ for } 1 \leqq t < m \text{ (if } m > 1). \tag{32}$$

It is seen from (32), (27) and (1) that

$$(m+1)[\beta_{m+1}] \leqq [\alpha * \beta_m]; \tag{33}$$

cf. (14), (15). But (33) and (28) show that

$$[\beta_{m+1}] \leqq n[\alpha][\beta_m]/(m+1). \tag{34}$$

If an induction is applied on (34), it is seen from (29) that the proof of (31) is now complete.

Let $c = n[\alpha]$. Then, since n is the fixed dimension number, $0 \leqq c < \infty$, by (16). Hence, from (31),

$$\sum_{m=1}^{\infty} \beta_m \leqq \sum_{m=1}^{\infty} c^m/m! < \infty.$$

It follows therefore from (14), (15) and (25) that the series

$$\beta(t) = \sum_{m=1}^{\infty} \beta_m(t)$$

defines on the half-line (1) a function $\beta(t)$ satisfying

$$\beta < \infty \qquad\qquad (\text{and } \beta(1) = 0).$$

It is also seen that this $\beta(t)$ and the assignment (3) define on the half-line (4) a function $B(s)$, and that the latter is representable on the half-line (4) in the form (21). In fact, the convergence of the series (21) at every $s \geqq 0$ and the legitimacy of the term-by-term integration leading from (3) and (24) to (21) are clear from the last three formula lines.

Finally, the functions $B_m(s)$ occurring in (21) have been defined, after (21), by the method of successive approximations, which belong to the initial condition (20). Hence it is clear, from the standard argument in successive approximations, that the matrix $X(s)$ which is defined on the half-line (4) by (6) and (21) is a solution of (18). That the assignment (20) is actually satisfied by the $B(s)$ constructed, is seen by using for $B(s)$ its representation (3). In fact, since $[\beta] < \infty$ and $t \geqq 1 > 0$ in (3), the integral (3) tends exponentially to 0, as $s \rightarrow \infty$.

In order to complete the proof of the Theorem, all that remains to be ascertained is that the columns of the matrix (6) are linearly independent, i. e., that $\det X(s) \neq 0$ for every s on (4). Since (18) is known to imply that

$$\det X(s) = \det X(s_0) \cdot \exp \int_{s_0}^{s} \operatorname{trace} A(t)\, dt,$$

it is sufficient to show that $\det X(s) \neq 0$ for some s on (4). But this is clear from (20) and (7).

APPENDIX.

The Theorem, thus proved, has extensions in two obvious directions.

On the one hand, the lower end-point, $t = 1$, of the half-line (1) can be replaced by any positive number, $t = p > 0$. For, if the lower limit of integration in (2) and (3) is any $t = p > 0$, it can be reduced to $t = 1$ by a change, $t \rightarrow pt$, of the integration variable. It is true that the e^{-st} in both definite integrals then becomes replaced by e^{-spt}. But this reduces to e^{-st} by the substitution $s \rightarrow s/p$, which is admissible, since it transforms the half-line (4) into itself.

On the other hand, it is clear from the absolute convergence of both integrals (2), (3) on the closed half-line (4) that the latter can be replaced by the half-plane ($\sigma \geqq 0, -\infty < t < \infty$) of a complex variable, $s = \sigma + it$. In particular, since this half-plane is closed, the Theorem is applicable on the boundary line, $\sigma = 0$, where $s = it$.

These two obvious extensions of the Theorem contain (and are substantially equivalent to) the following statement:

Let $A(t)$, where $-\infty < t < \infty$, be a matrix function of n by n elements each of which is representable in the form

$$\int_{p}^{\infty} e^{it\lambda} d\alpha_{jk}(\lambda), \qquad (j, k = 1, \cdots, n; -\infty < t < \infty),$$

where p is a positive number and each of the n^2 scalar functions $\alpha_{ij}(\lambda)$ has *a finite total variation on the half-line* $p \leqq \lambda < \infty$. *Then every vector* $x(t)$ *representing a solution of* $dx/dt = A(t)x$ *is of the form*

$$x(t) = c + \int_p^\infty e^{it\lambda} d\gamma(\lambda), \qquad -\infty < t < \infty,$$

where the vector c *is an integration constant, which can be assigned arbitrarily, and each of the* n *components of the vector* $\gamma(\lambda) = \gamma_c(\lambda)$ *has a finite total variation on the half-line* $p \leqq \lambda < \infty$.

If all n functions $\alpha_{jk}(\lambda)$ are chosen to be step-functions (which can have clustering points of discontinuity), there results the following corollary:

Let $A(t)$, *where* $-\infty < t < \infty$, *be a matrix function of* n *by* n *elements each of which is uniformly almost periodic. Suppose that*

(i) *the frequencies,* λ_m, *occurring in the Fourier expansion,*

$$A(t) \sim (\sum_m a_{jk}^{(m)} e^{i\lambda_m t}),$$

of $A(t)$ *have a positive lower bound, (or a negative upper bound), and that*

(ii) *these* n^2 *Fourier expansions are absolutely convergent,*

$$\sum_{j=1}^n \sum_{k=1}^n \sum_m | a_{jk}^{(m)} | < \infty.$$

Then every vector $x(t)$ *representing a solution of* $dx/dt = A(t)$

(I) *is uniformly almost periodic, with a mean-value,*

$$\mathfrak{M}(x) = \lim_{v-u\to\infty} (v-u)^{-1} \int_u^v x(t)\,dt,$$

which can be assigned as an arbitrary integration constant, $c = \mathfrak{M}(x)$, *determining* $x(t)$, *and*

(II) *the Fourier expansion of each of the* n *components of every* $x(t)$ *is absolutely convergent.*

In addition, every ray containing the frequencies λ_m *contains every non-vanishing frequency of each of the* n *components of every solution* $x(t)$.

The last assertion, in which the "ray" is meant to be generated by positive integral coefficients, corresponds to the Remark following the Theorem.

It would be interesting to know whether the last italicized theorem remains true if its assumption (ii) *and* its assertion (II) are omitted.

THE JOHNS HOPKINS UNIVERSITY.

ON 1-REGULAR CONVERGENCE OF SEQUENCES OF 2-MANIFOLDS.*

By GAIL S. YOUNG, JR.

By results of Whyburn's [5],[1] if a sequence of 2-spheres, in a compact space, converges 1-regularly,[2] then the limiting set is either a point or a 2-sphere; and if a sequence of 2-cells converges 1-regularly while the boundaries of the cells converge 0-regularly, then the limiting set is either a point or a 2-cell. Generalizations of these have been made by Vaughan [3], White [4] and Begle [1]. Of particular interest to us here is this result[3] proved by Begle: If a sequence of compact orientable 2-manifolds (without boundary) converges 1-regularly to a non-degenerate set P, then (1) P is itself a 2-manifold, and (2) P is homeomorphic to all but a finite number of terms of the sequence.

In this paper I intend to extend the above theorems to the case of 1-regular convergence of 2-manifolds in general.[4] If we try to get Begle's conclusion in the non-compact case, however, we find that the usual definition of 1-regular convergence is stronger than needed for (1) and not powerful enough for (2). To see this last statement consider the following example: Let M_0 denote the set in 3-space which is the sum of all the points in the xy-plane exterior to or on the circles $(x \pm 2)^2 + y^2 = 1$ plus all the points with positive z-coordinates lying on the torus of revolution cutting $z = 0$ in these two circles. For each natural number n, let M_n denote the image of M_0 under a translation sending each point (x, y, z) of M_0 into the point $(x + n, y, z)$. Then, in the usual sense $\{M_n\}$ converges 1-regularly to $z = 0$, but no M_n is homeomorphic to this plane. An example illustrating a slightly different trouble is furnished by the sequence of half-planes $\{z = 1/n,$

---

* Received May 10, 1948; presented to the American Mathematical Society, December 27, 1946 and September 4, 1948.

[1] Numbers in brackets refer to the bibliography.

[2] In a compact space, a sequence $\{A_n\}$ of closed sets converges r-regularly to a set A provided that, first, $\{A_n\}$ converges to A, and, second, given $\epsilon > 0$, there exist numbers $\delta > 0$ and N such that for $0 \leqq i \leqq r$, and $n > N$, every i-cycle of A_n of diameter $< \delta$ bounds in A_n on a set of diameter $< \epsilon$.

[3] Theorem 7 of [1]. In Begle's paper, this appears as an application of more general results on convergence of sequences of generalized orientable manifolds.

[4] By a 2-manifold, I mean here a connected metric space, each point of which has a neighborhood whose closure is homeomorphic to a closed 2-cell. The set of all points not having neighborhoods homeomorphic to open 2-cells is the boundary of the manifold.

$x \geqq -n\}$, which also converges 1-regularly to $z = 0$. To remove these difficulties, I make the following definitions.

We suppose throughout this paper that all sets considered are imbedded in a locally compact complete metric space S. We use Vietoris cycles and homologies on compact carriers only.

DEFINITIONS: A sequence of closed sets $\{A_n\}$ converges W-r-regularly to a set A provided that (1) $\lim A_n = A$; (2) for each point p of A and each $\epsilon > 0$ there exist positive numbers δ and N such that for each two integers n and k, $n > N$ and $0 \leqq k \leqq r$, every k-cycle of A_n in the δ-neighborhood of p bounds on a subset of A_n in the ϵ-neighborhood of p. The sequence converges S-r-regularly to A provided that (1) given $\epsilon > 0$, there is an integer N such that, for $n > N$, each point of A_n is at distance $< \epsilon$ from A, and each point of A is at distance $< \epsilon$ from A_n; and (2) given $\epsilon > 0$, there exist positive numbers δ, N such that, for $n > N$ and $0 \leqq k \leqq r$, every k-cycle of A_n of diameter $< \delta$ bounds on a subset of A_n of diameter $< \epsilon$.

For compact spaces these definitions coincide with r-regularity. We can now state the principal results of this note.

THEOREM 2. *Let $\{M_n\}$ be a sequence of 2-manifolds converging* W-1-*regularly to a non-degenerate set M. Suppose that the (combinatorial) boundaries B_n of the sets M_n are either empty or converge* W-0-*regularly or have an empty upper limit. Then M is a 2-manifold.*

THEOREM 3. *Let $\{M_n\}$ be a sequence of 2-manifolds converging* S-1-*regularly to a 2-manifold M. Suppose that either the boundaries of M and the sets M_n are empty, or the boundaries B_n of the sets M_n converge* S-0-*regularly to the boundary of M. Then M is homeomorphic to all but a finite number of the sets M_n.*

We prove some preliminary results.

LEMMA 1. *Let the closed set B be the limit of a sequence $\{B_n\}$ of closed sets converging* W-0-*regularly to B. Then*

(1.1) *If $\{x_n; x_n \varepsilon B_n\}$ and $\{y_n; y_n \varepsilon B_n\}$ converge to the point x of B, then for almost all n, $x_n + y_n$ is contained in a component of B_n;*

(1.2) *If for each n C_n is a component of B_n, and* $\liminf C_n \neq 0$, *then* $\liminf C_n$ *and* $\limsup C_n$ *are each closed and open in B;*

(1.3) *Each component of B is open in B;*

(1.4) *If the convergence is* S-0-*regular, then for almost all n there is a one-to-one correspondence t_n between the components of B and those of B_n with the property that if C is a component of B, then* $\lim t_n(C) = C$.

Proof. Let ϵ be chosen so small that the 2ϵ-neighborhood U of x has a compact closure. For n large, diam $(x_n + y_n + x)$ is so small that every 0-cycle of B_n carried by $x_n + y_n$ bounds in $U \cdot B_n$; and any irreducible carrier of this homology is a connected subset of B_n containing $x_n + y_n$. This proves (1. 1).

If lim inf $C_n = H$, then H is closed. Let x be a point of H, and choose any $\epsilon > 0$; then there exist numbers δ, N, satisfying the conditions in the definition of W-0-regularity for x and ϵ. Let y be a point of B at distance $< \delta/3$ from x, and let $\{y_n; y_n \varepsilon B_n\}$ converge to y. Let $\{x_n; x_n \varepsilon C_n\}$ converge to x. There is a number N' such that, if $n > N'$, $d(y, y_n)$ and $d(x, x_n)$ are both less than $\delta/3$. For $n > N + N'$, any 0-cycle of B_n carried by $x_n + y_n$ bounds in B_n on a connected subset, of diameter $< \epsilon$. Hence y_n also lies in C_n, so that y is in lim inf C_n, proving that H is open. A similar argument, using a subsequence of $\{C_n\}$, proves that lim sup C_n is open, and so establishes 1. 2. I remark in passing that for non-compact sets, 1. 2 cannot be strengthened, though for compact sets we can actually prove that lim inf C_n = lim sup C_n, using 1. 3.

Since every component, C, of B intersects lim inf C_n for some sequence $\{C_n\}$ of components of $\{B_n\}$, it follows from 1. 2 that C is in lim inf C_n. Let x be any point of C and choose ϵ, δ, N, y, $\{x_n\}$, $\{y_n\}$ and N' as in the previous paragraph, ϵ being so small that the 2ϵ-neighborhood U of x has a compact closure. Then if K_n denotes a subset of B_n of diameter $< \epsilon$, $n > N + N'$, on which $x_n - y_n$ irreducibly bounds, K_n is closed and connected, and so is lim sup K_n, from the compactness of $\bar{U}$. Hence y belongs to C, and 1. 3 follows.

If the convergence is S-0-regular, the argument used in the previous paragraph shows that there is a $\delta > 0$ such that if two points of B are at distance $< \delta$, then they lie in the same component of B. We may choose δ so that, similarly, there is an N, the same for all points of B, such that if x_n and y_n are points of B_n, $n > N$, which are within δ of a point x of B, then they lie in the same component of B_n. There is an N' such that each point of B_n, $n > N'$, is within $\delta/3$ of some point of B. For $n > N + N'$ if C is a component of B_n, then no point of C is within $\delta/3$ of each of two components of B, for if two such components existed, C would intersect the boundary of the $\delta/3$ neighborhood of their sum K, in some point. This is impossible, however, as then some third component of B has points within $2\delta/3$ of K. For $n > N + N'$, then, to each component of B_n make correspond the unique component of B at distance $< \delta/3$ from it. This satisfies the conditions of 1. 4.

THEOREM 1.[5] *Let* $\{B_n\}$ *be a sequence of closed sets converging* W-0-*regularly to a closed set* B. *If each component of each* B_n *is a simple closed curve or an open curve, and no component of* B *is degenerate, then each component of* B *is also a simple closed curve or an open curve. If the convergence is* S-0-*regular, then* B *is homeomorphic to almost all* $\{B_n\}$.

Proof. If each point of B is of Menger order 2, the first part of the theorem is true. Suppose, then, that some point x of a component C of B is of order other then 2. If order $x = 1$, there is a neighborhood U of x, with compact closure, such that $\beta(U) \cdot C$ is a point, $\beta(U)$ denoting the set-theoretic boundary of U, and such that $(S - \bar{U}) \cdot C \neq 0$. Then if for each n C_n is a component of B_n, and $\liminf C_n$ contains C, we can choose two sequences $\{x_n; x_n \varepsilon C_n\}$ and $\{z_n; z_n \varepsilon C_n\}$, the first converging to x, the second converging to a point z of C not in $\bar{U}$. One, two, or possibly three components of $C_n - x_n - z_n$ meet $\beta(U)$; in each such component an arc irreducibly contains its intersection with $\beta(U)$. Let H_n denote the sum of these arcs. If a_n and b_n are points of $\beta(U)$ in different components of H_n, then no 0-cycle carried by $a_n + b_n$ bounds in $C_n - x_n - z_n$. Since $C_n \cdot \beta(U)$ converges to y, for n large every 0-cycle carried by $a_n + b_n$ does bound in $C_n - x_n - z_n$, so that for n large H_n has just one component. Then H_n clearly separates C_n, one component of $C_n - H_n$ being in U. Since an arc cannot separate a simple closed curve, C_n is an open curve, which contradicts the compactness of $\bar{U}$.

Suppose that order $x > 2$. Then x is the junction point of a simple triod $ax + bx + cx$, since by Theorem 2.2 of [5] and Lemma 1, each component of B is locally connected. By Theorem 2.21 of [5], whose proof is still valid here, there exist sequences $\{A_{1n}\}$, $\{A_{2n}\}$ and $\{A_{3n}\}$ of arcs of C_n converging 0-regularly to axb, axc, and bxc, respectively. Since each C_n is atriodic, these can be chosen to be non-intersecting. In each A_{in}, select a point x_{in} such that $\lim x_{in} = x$. Then no 0-cycle carried by, say, $x_{1n} + x_{2n}$, bounds in $C_n - \beta(A_{1n} + A_{2n} + A_{3n})$. For n large, this contradicts W-0-regularity and proves the first part of the theorem.

In view of 1.4 of Lemma 1, all that is needed to finish the proof of Theorem 1 is to show that no sequence of simple closed curves can converge S-0-regularly to an open curve and that no sequence of open curves can converge even W-0-regularly to a simple closed curve. Suppose, then, that the sequence $\{J_n\}$ of simple closed curves converges S-0-regularly to the open curve L. In L there is a countable closed set H having no limit point and with the property that each point of L is between some two points of H (which corresponds to the set of all integers in the number line). No infinite subset of H forms a Cauchy sequence, since A is complete, so that there is an $\epsilon > 0$ such that no infinite subset of H is of diameter $\leqq 2\epsilon$. There is an N such that each set J_n, $n > N$, is within $\epsilon/2$ of L and such that each point of L is within $\epsilon/2$ of each such J_n. Let $\{x_i\}$ be an ordering of H. For each integer i, let y_i denote a point of J_n, $n > N$, at distance $< \epsilon/2$ from x_i. Some point y of J_n is a limit point of $\sum y_i$. Then infinitely many points x_i are within ϵ of y and hence form a subset of H of diameter $< 2\epsilon$, which is impossible.

---

[5] A more general theorem concerning 0-regular convergence of certain 1-dimensional sets could be proved, but there seems no great necessity for it. If in Theorem 1, each component of each B_n were a simple closed curve, Theorem 1 would follow from Lemma 1 and Theorem 3.2 of [5].

From the local compactness of our space, every simple closed curve has a neighborhood with compact closure. As an open curve is required to be closed, such a neighborhood cannot contain one, which would dispose of the second case. However, this argument rather avoids the issue, and it is true that the assumption of closure is not needed. Suppose, in fact, that a sequence $\{O_n\}$ of homeomorphs of the line, not necessarily closed, converged W-0-regularly to a simple closed curve $J = axb + ayb$. Examination of the proof of Theorem 2. 21 of [5] shows that in each O_n there exist arcs $a_nx_nb_n$, $a_ny_nb_n$, such that $\{a_nx_nb_n\}$ and $\{a_ny_nb_n\}$ converge 0-regularly to axb and ayb, respectively. For n large, this implies the existence of a simple closed curve in O_n, which is impossible.

The circles $x^2 + y^2 - 2ny = 0$ converge W-0-regularly to $y = 0$, showing that S-0-regularity was needed in Theorem 1.

Proof of Theorem 2. I have proved in [8] that a 2-manifold, N, is characterized among locally compact, locally connected and connected metric spaces by these two conditions: (1) N has no local cut points, and (2) if C is a compact subset of N, there is an $\epsilon > 0$ such that every simple closed curve in C of diameter $< \epsilon$ separates N. Hence we need only show that M satisfies these conditions. However, in the proof of his theorem mentioned above, Begle has given an argument which needs only slight modifications to prove that M satisfies (2), so that a proof is needed only for (1). Condition (1), however, is not trivial, and the reader may perhaps follow the rest of the argument better if he keeps in mind the following two examples. The first is formed by the sequence of sets $1 - 1/n \leqq x^2 + y^2 \leqq 1 + 1/n$, which converges 1-regularly to the unit circle, though the boundary does not converge 0-regularly. The second is formed by letting M denote a pinched torus and $\{M_n\}$ denote a sequence of tori converging smoothly to M, the convergence being 0-regular, and failing to be 1-regular only at the local cut point of M.

Suppose, then, that some point, x, of M were a local cut point of M. There is a neighborhood U of x with compact closure such that (1) letting $U_n = M_n \cdot U$, for almost all n every cycle of M_n that is carried by a subset of U_n bounds in M_n; (2) $M \cdot U$ is connected, but some two points, a and b, are separated in $M \cdot \bar{U}$ by x; (3) for each n, $U_n - B_n \cdot U$ contains two points, a_n and b_n, for which $\lim a_n = a$, $\lim b_n = b$; (4) a, b, and x are so close that for n large any 0-cycle of M_n carried by $a_n + b_n$ bounds in U_n, so that a_n and b_n lie in the same component of U_n; and (5) for no n does U_n contain all of a component of B_n, nor does it contain points of two components of B_n. The second part of (5) follows from Lemma 1. The first part of (5) follows from the fact that if it were not so, we could find a sequence of components of B_n containing a subsequence converging to x; since a cycle carried by a component of B_n can bound only on all M_n, if at all, W-1-regularity would imply $\liminf M_n = x$, an impossibility. Let V be a neighborhood of x whose closure lies in $U - a - b$. Then, for n large, $\bar{V} \cdot M_n$ separates a_n from b_n in U_n. Otherwise, for infinitely many n, there is an arc $a_n b_n$ in $U_n - M_n \cdot \bar{V}$, and $\limsup a_n b_n$ is a connected subset of $M \cdot (\bar{U} - V)$ that contains a and b but not x. There is no loss in assuming that V is so small that for some open neighborhood R of $\bar{V}$, and for n large, every 0- or 1-cycle of $R \cdot M_n$ bounds in $U_n - a_n - b_n$, and every 0-cycle of $B_n \cdot R$ bounds in $B_n \cdot U_n - a_n - b_n$. Choose one such value of n. By considering a triangulation of M_n of small enough norm we can obtain a closed set C_n which is the sum of a finite number of cells of the triangulation and which lies in $R \cdot M_n$ and contains $\bar{V} \cdot M_n$ in its interior—possibly C_n is not connected. There is no difficulty in assuming that each component of the closure of the mod 2 boundary of C_n (considered as a complex) is a simple closed curve,[6] though the mod 2 boundary of C_n may be larger than the set-theoretic boundary if $C_n \cdot B_n \neq 0$. Now I assert that one such simple closed curve separates a_n from b_n in U_n. For suppose not, and let $a_n b_n$ be an arc in U_n. Let J be the first component of the closure of the mod 2 boundary of C_n that $a_n b_n$ intersects. Since any cycle carried by J bounds in $U_n - a_n - b_n$ there is a connected set D_n in $U_n - a_n - b_n$, open rel. M_n, whose boundary rel. M_n is J. Let p and q be the first and last points of $a_n b_n \cdot J$ in order a_n to b_n in $a_n b_n$. Then neither $a_n p$ nor $a_n q$ is in D_n. By definition of J, the component of $U_n - J$ containing $a_n p - p$ cannot lie in C_n. But every point of J is a limit point of C_n, so that D_n must intersect C_n, and, indeed, every point of D_n near enough to J is in C_n. Hence the points of qb_n near q lie in $U_n - C_n$, J being in the closure of the mod 2 boundary of C_n. Now $a_n + b_n$ lies in one component, K, of $U_n - J$. Unless J intersects B_n, K has all of J for boundary, and even if J intersects B_n in more than one point, K has either all of J or an arc of J for boundary, a statement that requires some argument. First, by the condition that R is so small that every 0-cycle of $R \cdot B_n$ bounds in $B_n \cdot U$, and the fact that U is so small that no component of B_n lies entirely in U, we can find a unique arc T of $B_n \cdot U$ that irreducibly contains $J \cdot B_n$. Exactly one component, C, of $U_n - J$ both lies in U_n and has all of J for boundary. If L is a component of $J - B_n \cdot J$ then L spans an arc A of T, and $L + A$ is a simple closed curve separating U_n into exactly two connected domains, one of which contains C, and the other of which has all of $L + A$ for boundary, and is the only one of the two that does. Either K is C or is one of these components of $U_n - J$ whose boundary is such an arc L; in either case, our remark is proved. Using this result, then, it is possible to find an arc from a_np to qb_n in K which is close enough to J not to intersect C_n, thus constructing an arc a_nb_n not intersecting J at all. Repetition of this argument gives finally an arc a_nb_n not intersecting C_n but lying in U_n. Hence there is a simple closed curve, J_n, in the mod 2 boundary of C_n that separates a_n from b_n in U_n. Even more than this, some arc A_n of J_n separates a_n from b_n in U_n, for if J_n irreducibly separates U_n, it separates it into two connected sets, one of which lies in $U_n - a_n - b_n$, since any cycle carried by J_n bounds in that set. Of course, the existence of A_n implies that $J_n \cdot B_n \neq 0$, so that, as before, there is an arc T_n of $B_n \cdot U$ such that $A_n + T_n$ is a simple closed curve which separates U_n into two connected sets, one containing a_n, the other b_n. It will be recalled that B_n does not contain a_n or b_n, so that T_n contains neither. We now have a contradiction, since $A_n + T_n$ cannot bound in $U_n - a_n - b_n$. Hence M has no local cut points, which completes the proof of Theorem 2.

---

[6] Cf. Theorem of Roberts and Steenrod [2], for example, which, while proved only for a manifold without boundary, can clearly be extended.

Proof of Theorem 3. We shall form a particularly desirable triangulation of M. There exist numbers $\eta > 0$ and N_1 such that, for $n > N_1$, every 0- or 1-cycle of M_n (B_n) of diameter $< 3\eta$ bounds on a subset of M_n (B_n) of diameter $< 1/3 \operatorname{diam} M_n$ (B_n). (There are no small non-zero 1-cycles in B_n.) As may be seen by an argument similar to that of Lemma 1.1 of [5], there exist numbers γ, γ' such that every 0-cycle of M (B) of diameter $< \gamma'$ bounds on a subset of M (B) of diameter $< \gamma/3$, and such that every 1-cycle of M of diameter $< \gamma$ bounds on a subset of M of diameter $< \eta$. In M there is a set of points $\{v_i\}$ such that (1) every point of M is within $\gamma'/2$ of some

v_i; (2) every point of B is within $\gamma'/2$ of some v_i in B; and (3) no two points v_i are closer than $\gamma'/4$. These will be the vertices of our triangulation, Δ, with a 1-dimensional skeleton L. If two points v_i, v_j are in B, with $d(v_i, v_j) < \gamma'$, and if no other v_k is on the unique arc v_iv_j of B with diameter $< \gamma/3$, then put v_iv_j in L. If no more than one of v_i, v_j is in B, we can select an arc v_iv_j of diameter $< \gamma/3$, having at most one end point in common with B. By well known theorems on 2-manifolds, we can select all these arcs so that each two have no more in common than an end point. Put all these in L. Now let $J = v_1v_2 + v_2v_3 + v_1v_3$ be a simple closed curve in L. Then diam $J < \gamma$, so that any 1-cycle carried by J bounds on a subset of M of diameter $< \eta$. Using this, and the definition of L, we see that J separates M into two connected open sets, one of which, D, has diameter $< \eta$ and has all J for boundary. With the help of Theorem 3.1 of my paper [8] and the properties of γ, it follows that $D + J$ is a closed 2-cell. It is then easy to see that $L \cdot (D + J)$ defines a triangulation of $D + J$. It follows readily that L is the 1-dimensional skeleton of a triangulation Δ of M of norm $< \eta$, and that every 1- or 2-cell of Δ has diameter $> \gamma'/4$.

As has been said before, Whyburn has proved in Theorem 2.21 of [5] that every arc A of M is the limit of a sequence of arcs, one from each M_n, that converges 0-regularly to A. Examination of this proof will show that, by the aid of uniformity properties of S-1-regular convergence, we can select for each closed 1-cell, C, of L a sequence $\{C_n\}$ of arcs converging 0-regularly to C, such that (1) for each n, C_n is in M_n; (2) if C is in B, the boundary of M, then C_n is in B_n; and (3) given $\epsilon > 0$, there is an N such that for each such 1-cell, C, and $n > N$, each point of C_n is within ϵ of C, and each point of C is within ϵ of C_n. Since each 1-cell of L is of diameter $> \gamma'/4$, it follows from this and (3), that for n greater than some number N_2, the correspondence between the cells C and the arcs C_n is one-to-one. Hence, for each fixed $n > N_2$, the sum of the sets C_n, summed over the 1-cells of L, is an approximation, in some sense, to L; call this sum L'_n. I will modify L'_n to make it homeomorphic to L, for n large. First, it is clear—from Whyburn's argument, if nothing else—that we can assume that if $\{C_n\}$ and $\{C'_n\}$ are the sequences chosen to converge to 1-cells of L with a common vertex, then $C_n \cdot C'_n$ contains an end-point of each. Let x be a vertex of L, and let $y_1, \cdots, y_k$ be the vertices of L which can be joined to x by a 1-cell of L. For each j, $1 \leqq j \leqq k$, let $\{x_ny_{jn}\}$ denote the sequence of arcs chosen above to converge to xy_j, x_n converging to x. If for each n, $d_n(x)$ denotes the l. u. b. of distances $d(x_n, z)$, where z is any point common to two arcs x_ny_{jn}, then $d_n(x)$ converges to zero with n, uniformly in x, by (3) of our third sentence. There is an N_3 such that for $n > N_3$, and for any x,

$d_n(x) < 1/4 \min d(x_n, y_{jn})$, the minimum being with respect to j. For such an n, I will modify L_n at x so that the new arcs $x_n y_{jn}$ will have only x_n in common, by the following construction. If $x_n y_{1n} \cdot x_n y_{2n} \neq x_n$, then one of these, say $x_n y_{2n}$, is not entirely in B_n, since each point of B_n is of Menger order 2. Using standard accessibility and separation properties of 2-manifolds, we can construct an arc $x_n z_{2n}$, with z_{2n} in $x_n y_{2n}$, of diameter $< 1/4\ d_n(x)$, that has just x_n in common with $x_n y_{1n} + B_n$, and such that $y_{2n} z_{2n} + z_{2n} x_{2n}$ is an arc. Suppose now that for $1 \leqq i \leqq j-1 < k$, arcs $x_n z_{in}$ have been constructed so that $\operatorname{diam} x_n z_{in} < 3/2\ d_n(x)$, no two have more than x_n in common, and so that no one intersects B_n unless it lies therein. Then if $x_n y_{jn}$ is not in B_n, we can construct an arc $z_{jn} x_n$ from a point z_{jn} in $x_n y_{jn}$ to x_n, of diameter $< 3/2\ d_n(x)$ and having only x_n in common with $\sum_{i=1}^{j-1} x_n z_{in} y_{in} + B_n$. (Roughly, what we do is start from z_{jn}, follow $x_n y_{jn}$ along until we meet either the first modified arc or B_n, then run along next to that arc to x_n.) Of course, if $x_n y_{jn}$ is in B_n, nothing need be done to it for the induction. Proceeding in this manner, we can therefore modify L'_n at x_n so that L'_n as thus modified is locally homeomorphic to L at x. But by the uniformity of S-1-regularity, this process could be carried out simultaneously at every vertex of L'_n, yielding, for each n sufficiently large, 1-complexes L_n which are homeomorphic to L, under a homeomorphism sending each vertex of L_n into the nearest vertex of L. The sequence $\{L_n\}$ can easily be shown to converge S-0-regularly to L.

By the definition of η, for $n > N_1$, if x_1x_2, x_2x_3, and x_1x_3 are the (closed) 1-cells of the boundary of a 2-simplex of the triangulation Δ, then the subset $T = x_{1n}x_{2n} + x_{2n}x_{3n} + x_{1n}x_{3n}$ of L_n bounds an open set D in M_n, with $T + D$ compact. The set $T + D$ cannot contain a component of B_n, for such a component can bound only on all of M_n, if then, and this contradicts the definition of η, for n large enough. But, this being so, it is clear that $T + D$ has no local cut points. Further, every simple closed curve in $T + D$ distinct from T separates $T + D$. To see this, let J be such a simple closed curve and let x be a point of $T - T \cdot J$. There is an arc xy in $(M_n - T - D) + x$ from x to a point at least $1/2 \operatorname{diam} M_n$ from J. This arc does not intersect J, and since J bounds an open set E of diameter $< 1/2 \operatorname{diam} M_n$, $E \cdot (T + D)$ and $(M_n - \bar{E}) \cdot (T + D)$ both exist. Hence by Theorem 3.1 of my paper [8], $T + D$ is a 2-cell.

Now clearly the homeomorphism between L and L_n can be extended to one between M and a subset K of M_n. We need only prove that every point of M_n is in K. By Theorem 1, and the definition of L, every point of B_n is in L_n. Then K is an open subset of M_n, since the star of every point of K

in a 0- or 1-cell of $K - B_n$ is homeomorphic to an open subset of $M - B$. On the other hand, the diameters of the closed 2-cells forming K are bounded away from 0, by the definition of Δ, so that K is closed. Hence K is M.

I will point out, in passing, that just as Theorem 1 gives us information about 0-regular convergence of sets whose components are 1-manifolds, so we can prove a theorem on 1-regular convergence of sets whose components are 2-manifolds. The nature of such a theorem is so obvious that I will not state it.

Whyburn has proved that a sequence of 2-spheres (or 2-cells) converging 0-regularly converges to a cactoid (or hemi-cactoid), in other words to a set that can be obtained from the original sets by a monotone transformation. This is not a property of 2-manifolds in general. In other words, it is false that a sequence of compact 2-manifolds, converging 0-regularly, converges to a set that can be obtained by a monotone transformation from any compact 2-manifold whatever. Consider, for example, the set M which is the sum of the intervals $(0, 1)$ of the x- and y-axes, and of each of the intervals joining $(1/n, 0)$ to $(0, 1/n)$. For almost all null sequences $\{\epsilon_n\}$, the set M_n of all points ϵ_n away from M is a 2-manifold, and the sequences M_n converge 0-regularly to M. However, M has infinite coherence, whereas a monotone transformation can only lower coherence. It seems probable that a sequence $\{M_n\}$ of homeomorphic manifolds converging 0-regularly converges to a monotone image of its elements, but I have not tried to prove this. For further discussion of the 0-regular convergence problem, see Whyburn's paper [6].

UNIVERSITY OF MICHIGAN.

BIBLIOGRAPHY.

1. E. G. Begle, "Regular convergence," *Duke Mathematical Journal*, vol. 11 (1944), pp. 441-450.
2. J. H. Roberts and N. E. Steenrod, "Monotone transformations of two-dimensional manifolds," *Annals of Mathematics*, vol. 39 (1938), pp. 851-862.
3. H. E. Vaughan, Abstract no. 217, *Bulletin of the American Mathematical Society*, vol. 42 (1936), p. 337.
4. P. A. White, "On r-regular convergence," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 123-128.
5. G. T. Whyburn, "On sequences and limiting sets," *Fundamenta Mathematicae*, vol. 25 (1935), pp. 408-426.
6. ———, "Regular convergence and monotone transformations," *American Journal of Mathematics*, vol. 57 (1935), pp. 902-906.
7. ———, *Analytic Topology*, American Mathematical Society Colloquium Publications, vol. 28 (1942).
8. G. S. Young, "A characterization of 2-manifolds," *Duke Mathematical Journal*, vol. 14 (1947), pp. 979-990.

ON THE APPROXIMATION OF IRRATIONAL NUMBERS BY THE CONVERGENTS OF THEIR CONTINUED FRACTIONS.[*]

By Alfred Brauer and Nathaniel Macon.

Introduction. Let ξ be any positive irrational number. We shall consider the expansion of ξ as the regular continued fraction

$$\xi = [q_0, q_1, q_2, \cdots].$$

Let us denote the numerators and denominators of its convergents by A_n and B_n respectively $(n = 0, 1, 2, \cdots)$, and set

$$(1) \qquad | \xi - A_n/B_n | = 1/(\lambda_n B_n^2).$$

Hurwitz [10] proved that for any given ξ there exist infinitely many $\lambda_n > \sqrt{5}$, and that there exist numbers ξ for which only a finite number of the λ_n are greater than c for any $c > \sqrt{5}$. Vahlen [16] showed that for every ξ at least one of two consecutive λ_n exceeds 2, and Borel [1] proved that

$$\max(\lambda_n, \lambda_{n+1}, \lambda_{n+2}) > \sqrt{5},$$

for every n. These results were improved for special cases in which conditions were imposed upon the partial quotients q_n.

Let us assume that $q_{n+2} \geqq 2$. It was shown by Humbert [9] that

$$\max(\lambda_n, \lambda_{n+1}, \lambda_{n+2}) > \sqrt{8}, \qquad (n = 0, 1, 2, \cdots),$$

and by Fujiwara [3] that either $\lambda_{n+1} > 2.5$ or

$$\min(\lambda_n, \lambda_{n+2}) > 2.5.$$

Moreover, the following theorems were obtained by Fujiwara [3, 4]:

If $q_{n+2} = 2$ and $q_{n+1} = 1$, then at least one of the five numbers $\lambda_n, \lambda_{n+1}, \lambda_{n+2}, \lambda_{n+3}, \lambda_{n+4}$ is greater than $(2 + 5\sqrt{10})/6$.

If ξ is equivalent to neither $\frac{1}{2}(1 + \sqrt{5})$ nor $1 + \sqrt{2}$, if $q_{n+2} = 2$ and $q_{n+3} = 1$, and all the q_i are less than 3, then

$$\max(\lambda_n, \lambda_{n+1}, \lambda_{n+2}, \lambda_{n+3}, \lambda_{n+4}) > \sqrt{221}/5,$$

except for

$$[q_{n+2}, q_{n+3}, q_{n+4}, \cdots, q_{n+9}] = [2, 1, 1, 2, 2, 1, 1, 2].$$

---

[*] Received May 27, 1948. Presented to the American Mathematical Society, December 28, 1948.

Other proofs of these results were given by Perron [13], Hardy and Wright [7], L. R. Ford [2], Kurosa [12], Fukasawa (Morimoto) [5, 6], and Shibata [15]. For a more detailed history of the problem see Koksma [11].

In this paper, we shall prove the following extensions of Borel's theorem:

For every irrational number ξ, either at least two of each set of five consecutive λ_n exceed $\sqrt{5}$, or at least one of them exceeds 3.

Let t be any number of form $3m+2$. In each sequence of t consecutive λ_n, there are either at least $m+1$ elements which exceed $\sqrt{5}$ or m elements which exceed 3.

These results will be improved further for the case in which all of the partial quotients are less than 3. Moreover, we shall consider some further relations between consecutive numbers λ_n. In particular, we shall prove that for every n

$$\min(\lambda_n, \lambda_{n+2}) > \lambda_{n+1}/(\lambda_{n+1}-1),$$

and

$$\lambda_n\lambda_{n+1} > \lambda_n + \lambda_{n+1} > \max\{\lambda_n^2/(\lambda_n - 1), \lambda_{n+1}^2/(\lambda_{n+1}-1)\} > 4.$$

Heawood [8] and Perron [14] investigated those numbers ξ for which the upper limit of the λ_n is less than 3. It follows here that for each such number, only a finite number of λ_n may be less than 3/2.

Finally, estimates for the sums of three, four, and eight consecutive λ_n are obtained. It follows from these results that

$$\underline{\lim}\,(\sum_{i=0}^{m-1}\lambda_i)/m \geqq (31+15\sqrt{5})/32 > 2.0169.$$

1. The theorems of Vahlen and Borel and some consequences.

We will need the following well-known facts about continued fractions which may be found in Hardy and Wright [7], p. 150 ff. or Perron [13], p. 48 ff.

Let us denote the m-th complete quotient, $[q_m, q_{m+1}, \cdots]$ of the continued fraction ξ by ξ_m and set

$$\phi_1 = 0, \text{ and } \phi_m = B_{m-2}/B_{m-1}, \qquad (m = 2, 3, \cdots). \tag{2}$$

Then we have

$$|\xi - A_m/B_m| = 1/B_m^2(\xi_{m+1} + \phi_{m+1}) = 1/B_m^2(\xi_{m+2}^{-1} + \phi_{m+2}^{-1}).$$

Hence

$$\lambda_m = \xi_{m+1} + \phi_{m+1} = \xi_{m+2}^{-1} + \phi_{m+2}^{-1}. \tag{3}$$

Moreover, we have

$$|\xi - A_m/B_m| + |\xi - A_{m+1}/B_{m+1}| = 1/B_mB_{m+1}. \tag{4}$$

Two consecutive values ϕ_m and ϕ_{m+1} satisfy the condition

$$\min(\phi_m, \phi_{m+1}) < \tfrac{1}{2}(\sqrt{5}-1). \tag{5}$$

Finally, if we denote the partial quotients by q_m $(m = 0, 1, 2, \cdots)$, then

$$\xi_m = q_m + \xi_{m+1}^{-1}, \qquad (m = 0, 1, 2, \cdots), \tag{6}$$

$$\phi_{m+1} = (q_m + \phi_m)^{-1}, \qquad (m = 1, 2, \cdots). \tag{7}$$

Let us now prove the following lemma, which is due in part to Fujiwara [3]:

Lemma. *For all n, we have*

$$\phi_{n+2} = \tfrac{1}{2}\lambda_{n+1}\{1 - (1 - 4/\lambda_n\lambda_{n+1})^{\frac{1}{2}}\}, \tag{8}$$

$$\xi_{n+2} = \tfrac{1}{2}\lambda_{n+1}\{1 + (1 - 4/\lambda_n\lambda_{n+1})^{\frac{1}{2}}\}. \tag{9}$$

Proof. It follows from (1) and (4) that

$$1/\lambda_n B_n^2 + 1/\lambda_{n+1}B_{n+1}^2 = |\xi - A_n/B_n| + |\xi - A_{n+1}/B_{n+1}| = 1/B_nB_{n+1};$$

and, multiplying by $\lambda_{n+1}B_n^2$ and applying (2), we have

$$\phi_{n+2}^2 - \lambda_{n+1}\phi_{n+2} + \lambda_{n+1}/\lambda_n = 0. \tag{10}$$

Moreover, by (3)

$$\lambda_{n+1} = \xi_{n+2} + \phi_{n+2},$$
$$\lambda_n = \xi_{n+2}^{-1} + \phi_{n+2}^{-1}.$$

Hence,

$$\xi_{n+2}\phi_{n+2} = \lambda_{n+1}/\lambda_n. \tag{11}$$

If we consider (10) as a quadratic equation for ϕ_{n+2}, it follows from (3) and (11) that ξ_{n+2} and ϕ_{n+2} are the roots; and since $\xi_{n+2} > 1 > \phi_{n+2}$, the lemma is proved.

Fujiwara obtained from (8) the following proof of Vahlen's Theorem: since ϕ_{n+2} is real, we have

$$\max(\lambda_n, \lambda_{n+1}) > 2.$$

Now we shall prove the following somewhat modified version of Borel's Theorem:

Theorem 1. *If*

$$\phi_{n+2} < \tfrac{1}{2}(\sqrt{5}-1) \tag{12}$$

and $\lambda_n \leqq \sqrt{5}$, *then*

$$\lambda_n + \lambda_{n+1} > 2\sqrt{5},$$

that is, under the assumption (12), *either* $\lambda_n > \sqrt{5}$ *or* $\lambda_{n+1} > \sqrt{5}$.

It is seen from (5) that Theorem 1 contains Borel's Theorem.

Proof. The function $x + x^{-1}$ is decreasing for $x \leqq \frac{1}{2}(\sqrt{5} - 1)$. Thus it follows from (3) and (12) that

$$\lambda_n + \lambda_{n+1} = \xi_{n+2} + \xi_{n+2}^{-1} + \phi_{n+2} + \phi_{n+2}^{-1} > \xi_{n+2} + \xi_{n+2}^{-1} + \sqrt{5}. \tag{13}$$

Since $\lambda_n \leqq \sqrt{5}$, we have

$$\xi_{n+2}^{-1} + \phi_{n+2}^{-1} \leqq \sqrt{5},$$
$$\xi_{n+2}^{-1} \leqq \sqrt{5} - \phi_{n+2}^{-1} < \sqrt{5} - \tfrac{1}{2}(\sqrt{5} + 1) = \tfrac{1}{2}(\sqrt{5} - 1).$$

Hence, from (13),

$$\lambda_n + \lambda_{n+1} > 2\sqrt{5}.$$

THEOREM 2. *If* $\phi_{n+2} \geqq \frac{1}{2}$ *then* $q_{n+1} = 1$.

Proof. The theorem follows at once from (7), since $q_{n+1} + \phi_{n+1} \leqq 2$.

The following extension of Vahlen's theorem is now an immediate consequence:

THEOREM 3. *If every partial quotient is greater than* 1, *then*

$$\max(\lambda_n, \lambda_{n+1}) > \sqrt{5} \qquad (n = 0, 1, 2, \cdots). \tag{14}$$

Proof. By Theorem 2, we have $\phi_{n+2} < \frac{1}{2} < \frac{1}{2}(\sqrt{5} - 1)$ for every n; and so (14) follows from Theorem 1.

2. The sum and the product of two consecutive λ_n. It follows from (8) that $\lambda_n\lambda_{n+1} > 4$. More exactly, we have

THEOREM 4. *The following inequalities hold for every* n:

$$\lambda_n\lambda_{n+1} > \lambda_n + \lambda_{n+1} > \max\{\lambda_n^2/\lambda_n - 1), \lambda_{n+1}^2/(\lambda_{n+1} - 1)\} > 4.$$

Proof. Two cases arise.

Case I. $\lambda_{n+1} > 2$

Since $\phi_{n+2} < 1$, it follows from (8) that

$$\lambda_{n+1}(1 - 4/\lambda_n\lambda_{n+1})^{\frac{1}{2}} > \lambda_{n+1} - 2;$$

and so

$$\lambda_{n+1}^2 - 4\lambda_{n+1}/\lambda_n > \lambda_{n+1}^2 - 4\lambda_{n+1} + 4,$$
$$\lambda_n + \lambda_{n+1} < \lambda_n\lambda_{n+1}. \tag{15}$$

Hence,

$$(16) \qquad \lambda_n > \lambda_{n+1}/(\lambda_{n+1} - 1) \text{ and } \lambda_{n+1} > \lambda_n/(\lambda_n - 1).$$

Moreover, we obtain from (16)

$$(17) \qquad \lambda_n + \lambda_{n+1} > \lambda_{n+1}/(\lambda_{n+1} - 1) + \lambda_{n+1} = \lambda_{n+1}^2/(\lambda_{n+1} - 1) > 4,$$

since the function $x^2/(x - 1)$ assumes its minimum for $x = 2$, and, by (1), λ_{n+1} is irrational. Similarly,

$$(18) \qquad \lambda_n + \lambda_{n+1} > \lambda_n + \lambda_n/(\lambda_n - 1) = \lambda_n^2/(\lambda_n - 1) > 4.$$

Thus for Case I, the theorem follows from (15), (17), and (18).

Case II. $\lambda_{n+1} < 2$.

Since $\xi_{n+2} > 1$, we have, from (9),

$$\lambda_{n+1}(1 - 4/\lambda_n\lambda_{n+1})^{\frac{1}{2}} > 2 - \lambda_{n+1},$$

and the proof is exactly as in Case I.

From (16) we obtain at once

THEOREM 5. *For all n, we have*

$$\min(\lambda_n, \lambda_{n+2}) > \lambda_{n+1}/(\lambda_{n+1} - 1).$$

COROLLARY 1. *If only a finite number of the λ_n exceed 3, then only a finite number of them are less than 3/2.*

COROLLARY 2. *If $\lambda_n < \sqrt{5}$ and $\lambda_{n+1} < \sqrt{5}$, then*

$$\min(\lambda_n, \lambda_{n+1}) > (5 + \sqrt{5})/4.$$

Proof. This follows at once from the fact that $\lambda_n/(\lambda_n - 1)$ and $\lambda_{n+1}(\lambda_{n+1} - 1)$ are monotonic decreasing.

THEOREM 6. *If $\phi_{n+2} < \frac{1}{2}(\sqrt{5} - 1)$, then $\lambda_n\lambda_{n+1} > \lambda_n + \lambda_{n+1} > 2 + \sqrt{5}$.*

Proof. On putting $m = n + 1$ in (3), we have from, Theorem 4,

$$\begin{aligned}\lambda_n\lambda_{n+1} > \lambda_n + \lambda_{n+1} &= \xi_{n+2}^{-1} + \phi_{n+2}^{-1} + \xi_{n+2} + \phi_{n+2} \\ &> 2 + \phi_{n+2} + \phi_{n+2}^{-1} > 2 + \tfrac{1}{2}(\sqrt{5} - 1) + \tfrac{1}{2}(\sqrt{5} + 1).\end{aligned}$$

Hence

$$\lambda_n\lambda_{n+1} > \lambda_n + \lambda_{n+1} > 2 + \sqrt{5}.$$

COROLLARY. *For every n, we have either*

$$\lambda_{n+1}\lambda_{n+2} > \lambda_{n+1} + \lambda_{n+2} > 2 + \sqrt{5},$$

or both

$$\lambda_n\lambda_{n+1} > \lambda_n + \lambda_{n+1} > 2 + \sqrt{5} \text{ and } \lambda_{n+2}\lambda_{n+3} > \lambda_{n+2} + \lambda_{n+3} > 2 + \sqrt{5}.$$

For two important special cases, we have the following improvement of Theorem 6:

THEOREM 7. *If* $\phi_{n+2} < \frac{1}{2}(\sqrt{5} - 1)$ *and* $\lambda_{n+1} \geqq \sqrt{5}$, *then*

(19) $$\lambda_n + \lambda_{n+1} > 2\sqrt{5},$$

and

(20) $$\lambda_n \lambda_{n+1} > 5.$$

Proof. If $\lambda_n \leqq \sqrt{5}$, then (19) follows from Theorem 1. For $\lambda_n > \sqrt{5}$ (19) is obvious.

In order to prove (20), we note from (8) that

$$2\phi_{n+2} = \lambda_{n+1}\{1 - (1 - 4/\lambda_n\lambda_{n+1})^{\frac{1}{2}}\} < \sqrt{5} - 1;$$

and so

$$\lambda_{n+1} - \sqrt{5} + 1 < \lambda_{n+1}(1 - 4/\lambda_n\lambda_{n+1})^{\frac{1}{2}}.$$

The left member is positive, and so

$$\lambda_{n+1}^2 + 6 - 2\sqrt{5} - 2\lambda_{n+1}(\sqrt{5} - 1) < \lambda_{n+1}^2 - 4\lambda_{n+1}/\lambda_n,$$

$$\lambda_n\{\lambda_{n+1}(\sqrt{5} - 1) - (3 - \sqrt{5})\} > 2\lambda_{n+1}.$$

Since the coefficient of λ_n is positive, we have

$$\lambda_n > 2\lambda_{n+1}/\{(\sqrt{5} - 1)\lambda_{n+1} - 3 + \sqrt{5}\}.$$

Hence

(21) $$\lambda_n\lambda_{n+1} > 2\lambda_{n+1}^2/\{(\sqrt{5} - 1)\lambda_{n+1} - 3 + \sqrt{5}\}.$$

The right member is differentiable for

$$\lambda_{n+1} \neq (3 - \sqrt{5})/(\sqrt{5} - 1) = \tfrac{1}{2}(\sqrt{5} - 1)$$

and has minima at $\lambda_{n+1} = 0$ and $\lambda_{n+1} = \sqrt{5} - 1$. Thus it increases for $\lambda_{n+1} \geqq \sqrt{5}$; and, from (21), we obtain $\lambda_n\lambda_{n+1} > 5$.

COROLLARY. *If* $\lambda_{n-1} \leqq \sqrt{5}$ *and* $\lambda_n \leqq \sqrt{5}$, *then* $\lambda_n\lambda_{n+1} > 5$.

THEOREM 8. *If* $\phi_{n+2} < \frac{1}{2}(\sqrt{5} - 1)$ *and* $\phi_{n+3} < \frac{1}{2}(\sqrt{5} - 1)$, *then*

(22) $$\lambda_n + \lambda_{n+1} > 2{,}5 + \sqrt{5}$$

and

(23) $$\lambda_n\lambda_{n+1} > (11 + 5\sqrt{5})/4.$$

Proof. We have, from (7),

$$q_{n+2} + \phi_{n+2} = \phi_{n+3}^{-1} > \tfrac{1}{2}(\sqrt{5} + 1),$$

and so

$$q_{n+2} > \tfrac{1}{2}(\sqrt{5} + 1) - \tfrac{1}{2}(\sqrt{5} - 1) = 1.$$

Hence, by (6)

$$(24) \qquad \xi_{n+2} > q_{n+2} \geqq 2.$$

It follows, then, from (3) and (24) that

$$\begin{aligned}\lambda_n + \lambda_{n+1} &= \xi_{n+2}^{-1} + \phi_{n+2}^{-1} + \xi_{n+2} + \phi_{n+2}\\ &> 2.5 + \phi_{n+2}^{-1} + \phi_{n+2} > 2.5 + \sqrt{5}.\end{aligned}$$

Similarly, we have

$$\lambda_n\lambda_{n+1} = 2 + \phi_{n+2}/\xi_{n+2} + \xi_{n+2}/\phi_{n+2}.$$

But, by (24),

$$\xi_{n+2}/\phi_{n+2} > 4/(\sqrt{5} - 1) = \sqrt{5} + 1,$$

so that

$$\lambda_n\lambda_{n+1} > 2 + \sqrt{5} + 1 + (\sqrt{5} + 1)^{-1} = (11 + 5\sqrt{5})/4.$$

Remark. It follows from (24) and Theorem 2 that the assumptions of Theorem 8 are possible only if $\phi_{n+3} < \frac{1}{2}$.

3. The main theorems.

We are now able to prove

THEOREM 9. *In each sequence of five consecutive terms,* $\lambda_{n-2}, \lambda_{n-1}, \lambda_n, \lambda_{n+1}, \lambda_{n+2}$, *either at least two terms exceed* $\sqrt{5}$ *or* λ_n *is greater than* 3.

Proof. If only one of the five numbers is greater than $\sqrt{5}$, then, by Borel's Theorem, it must be λ_n; whence it follows from Theorem 1 that

$$(25) \qquad \min(\phi_n, \phi_{n+3}) > \tfrac{1}{2}(\sqrt{5} - 1).$$

Then, by (5),

$$(26) \qquad \max(\phi_{n-1}, \phi_{n+1}, \phi_{n+2}) < \tfrac{1}{2}(\sqrt{5} - 1),$$

thus, by (24),

$$(27) \qquad q_{n+1} \geqq 2;$$

and, by Theorem 2,

$$(28) \qquad q_{n-1} = q_{n+2} = 1.$$

Moreover, by (3),

$$\lambda_{n-1} = \xi_n + \phi_n < \sqrt{5},$$

and, by (25),

$$\xi_n < \sqrt{5} - \tfrac{1}{2}(\sqrt{5} - 1) = \tfrac{1}{2}(\sqrt{5} + 1).$$

Hence we have, from (6),

$$q_n = 1. \tag{29}$$

By repeated applications of (6) and (7), we may write ξ_{n+1} and ϕ_{n+1} as continued fractions and obtain, by (27), (28), and (26),

$$\begin{aligned} \lambda_n = \xi_{n+1} + \phi_{n+1} &= [q_{n+1}, q_{n+2}, \xi_{n+3}] + [0, q_n, q_{n-1} + \phi_{n-1}] \\ &> [2, 1, 1] + [0, 1, 1] = 3. \end{aligned} \tag{30}$$

In exactly the same way, we obtain the following result, of which Theorem 9 is a special case:

THEOREM 10. *Let t be any integer of form $3m + 2$. Each sequence of t consecutive λ_n contains either at least $m + 1$ terms which exceed $\sqrt{5}$, or at least m terms which are greater than* 3.

We shall improve Theorem 10 for a particularly interesting special case, namely that in which all the partial quotients q_n are less than 3. We understand by $[(\alpha, \beta, \gamma)_\rho]$ the finite periodic continued fraction $[\alpha, \beta, \gamma, \alpha, \beta, \gamma, \cdots]$, where the period occurs ρ times.

THEOREM 11. *Let ξ be an irrational number having only* 1*'s and* 2*'s in its expansion as a continued fraction and t be any integer of form $3m + 2$. In each sequence of t consecutive terms, $\lambda_{n-2}, \lambda_{n-1}, \lambda_n, \cdots, \lambda_{n+3m-1}$, either there are at least $m + 1$ terms which exceed $\sqrt{5}$, or the following relations hold for $\mu = 0, 1, 2, \cdots, m - 1$:*

if m and μ are both even, then

$$\lambda_{n+3\mu} > [(2, 1, 1)_{m-1-\mu}, 2, 1, 3] + [0, (1, 1, 2)_\mu, 1, 4/3];$$

if m is even and μ is odd, then

$$\lambda_{n+3\mu} > [(2, 1, 1)_{m-\mu}] + [0, (1, 1, 2)_\mu, 1, \tfrac{1}{2}(\sqrt{5} + 1)];$$

if m is odd and μ is even, then

$$\lambda_{n+3\mu} > [(2, 1, 1)_{m-\mu}] + [0, (1, 1, 2)_\mu, 1, 4/3];$$

and finally, if m and μ are both odd, then

$$\lambda_{n+3\mu} > [(2, 1, 1)_{m-1-\mu}, 2, 1, 3] + [0, (1, 1, 2)_\mu, 1, \tfrac{1}{2}(\sqrt{5} + 1)].$$

Proof. If only m of the terms exceed $\sqrt{5}$, then, as in the proof of Theorem 9, we have

$$q_{n+1+3\mu} = 2, \qquad (\mu = 0, 1, \cdots, m - 1),$$

$$q_{n+3\mu} = 1, \qquad (\mu = 0, 1, \cdots, m - 1),$$

and

$$q_{n-1+3\mu} = 1, \qquad\qquad (\mu = 0, 1, \cdots, m).$$

As in (30), we obtain

$$\xi_{n+1} = [2, 1, \xi_{n+3}] = [2, 1, 1, 2, 1, \xi_{n+6}] = \cdots$$
$$= [(2, 1, 1)_{m-1}, 2, 1, \xi_{n+3m}].$$

Since $3 > \xi_{n+3m} > 1$, by (6), we have

$$\xi_{n+1} > [(2, 1, 1)_{m-1}, 2, 1, 1] = [(2, 1, 1)_m], \text{ if } m \text{ is odd};$$

and

$$\xi_{n+1} > [(2, 1, 1)_{m-1}, 2, 1, 3], \text{ if } m \text{ is even}.$$

More generally, we obtain for $\mu = 0, 1, \cdots, m-1$,

$$(31) \qquad \xi_{n+1+3\mu} > [(2, 1, 1)_{m-\mu}] \text{ if } m-\mu \text{ is odd};$$

and

$$(32) \qquad \xi_{n+1+3\mu} > [(2, 1, 1)_{m-\mu-1}, 2, 1, 3] \text{ if } m-\mu \text{ is even}.$$

Similarly, we have for $\mu = 0, 1, \cdots, m-1$

$$(33) \qquad \phi_{n+1+3\mu} = [0, (1, 1, 2)_\mu, 1, q_{n-1} + \phi_{n-1}] > [0, (1, 1, 2)_\mu, 1, 4/3] \quad \text{if } \mu \text{ is even};$$

since $1/3 < \phi_{n-1} < \frac{1}{2}(\sqrt{5} - 1)$, by (7) and (26). Finally

$$(34) \qquad \phi_{n+1+3\mu} > [0, (1, 1, 2)_\mu, 1, \tfrac{1}{2}(\sqrt{5} + 1)] \text{ if } \mu \text{ is odd}.$$

The theorem follows at once from (31), (32), (33), and (34).

4. The sum of more than two consecutive λ_n.

THEOREM 12. *For every n, we have*

$$\lambda_n + \lambda_{n+1} + \lambda_{n+2} > \tfrac{1}{2}(5 + 3\sqrt{5}).$$

Proof. By Borel's Theorem, at least one of the three numbers $\lambda_n, \lambda_{n+1}, \lambda_{n+2}$ is greater than $\sqrt{5}$. First we assume that $\lambda_n > \sqrt{5}$. It follows from Theorem 4 that $\lambda_{n+1} + \lambda_{n+2} > 4$. Hence

$$(35) \qquad \lambda_n + \lambda_{n+1} + \lambda_{n+2} > \sqrt{5} + 4 > \tfrac{1}{2}(5 + 3\sqrt{5}).$$

If $\lambda_{n+2} > \sqrt{5}$, the proof is exactly the same. Thus we need to consider only the case in which $\lambda_{n+1} > \sqrt{5}$. By (5), at least one of the numbers ϕ_{n+2} and

ϕ_{n+3} is less than $\frac{1}{2}(\sqrt{5}-1)$. Let us first assume it is ϕ_{n+2}. It follows from Theorem 6 that

$$(36) \qquad \lambda_n > (2+\sqrt{5})/\lambda_{n+1},$$

and from Theorem 5 that

$$(37) \qquad \lambda_{n+2} > \lambda_{n+1}/(\lambda_{n+1}-1).$$

Hence, by (36) and (37),

$$(38) \qquad \lambda_n + \lambda_{n+1} + \lambda_{n+2} > (2+\sqrt{5})/\lambda_{n+1} + \lambda_{n+1} + \lambda_{n+1}/(\lambda_{n+1}-1).$$

Let us write λ for λ_{n+1} and denote the right member of (38) by $f(\lambda)$. The maxima and minima of $f(\lambda)$ are taken at the real roots of

$$g(\lambda) = \lambda^4 - 2\lambda^3 - (2+\sqrt{5})\lambda^2 + 2(2+\sqrt{5})\lambda - 2 - \sqrt{5} = 0.$$

We have

$$g'(\lambda) = 4\lambda^3 - 6\lambda^2 - 2(2+\sqrt{5})\lambda + 2(2+\sqrt{5}),$$
$$g''(\lambda) = 12\lambda^2 - 12\lambda - 2(2+\sqrt{5}).$$

It is sufficient to consider $\lambda > \sqrt{5}$, for $\lambda = \lambda_{n+1}$. Now $g''(\lambda) > 0$ for $\lambda > \sqrt{5}$ and so $g'(\lambda)$ is monotonic increasing for $\lambda > \sqrt{5}$. Since $g'(\sqrt{5}) > 0$, it follows that $g'(\lambda) > 0$ for $\lambda > \sqrt{5}$. Hence $g(\lambda)$ can have at most one root greater than $\sqrt{5}$. On the other hand, $\lambda = \frac{1}{2}(3+\sqrt{5})$ is a root of $g(\lambda) = 0$. Hence if $\lambda > \sqrt{5}$, $f(\lambda)$ has its only minimum at $\lambda = \frac{1}{2}(3+\sqrt{5})$. Thus it follows from (38) that

$$(39) \qquad \lambda_n + \lambda_{n+1} + \lambda_{n+2} > f\{\tfrac{1}{2}(3+\sqrt{5})\} = \tfrac{1}{2}(5+3\sqrt{5}).$$

Secondly, if $\phi_{n+3} < \frac{1}{2}(\sqrt{5}-1)$, then the left-hand sides of (36) and (37) must be interchanged. But (38) and the proof of (39) remain the same. Theorem 12 now follows from (35) and (39).

If $\lambda_{n+1} > \frac{1}{2}(3+\sqrt{5})$, we may use (37) instead of (36) for λ_n and λ_{n+2} and obtain

$$(40) \qquad \lambda_n + \lambda_{n+1} + \lambda_{n+2} > 2\lambda_{n+1}/(\lambda_{n+1}-1) + \lambda_{n+1},$$

which is sharper than (38). For $\lambda_{n+1} = \frac{1}{2}(3+\sqrt{5})$, (40) gives us the same estimate as does (39).

Now let us consider four consecutive λ_i.

THEOREM 13. *For every n, we have*

$$\lambda_n + \lambda_{n+1} + \lambda_{n+2} + \lambda_{n+3} > (21+5\sqrt{5})/4.$$

Proof. We must distinguish between two cases.

Case I. Only one of the four λ_i is greater than $\sqrt{5}$.

Clearly either $\lambda_{n+1} > \sqrt{5}$ or $\lambda_{n+2} > \sqrt{5}$.

(a) If $\lambda_{n+1} > \sqrt{5}$ then, by Theorem 4, we have

$$\lambda_{n+2} + \lambda_{n+3} > 4 \tag{41}$$

and

$$\lambda_n + \lambda_{n+1} > \lambda_{n+1}^2/(\lambda_{n+1} - 1) > 5/(\sqrt{5} - 1) = 5(\sqrt{5} + 1)/4. \tag{42}$$

Hence,

$$\lambda_n + \lambda_{n+1} + \lambda_{n+2} + \lambda_{n+3} > (21 + 5\sqrt{5})/4.$$

(b) If $\lambda_{n+2} > \sqrt{5}$, the proof can be obtained by interchanging the left-hand sides of (41) and (42).

Case II. At least two of the λ_i are greater than $\sqrt{5}$.

(a) $\max(\lambda_n, \lambda_{n+1}) > \sqrt{5}$, and $\max(\lambda_{n+2}, \lambda_{n+3}) > \sqrt{5}$.
From Theorem 4, we obtain

$$\lambda_n + \lambda_{n+1} > 5(\sqrt{5} + 1)/4,$$
$$\lambda_{n+2} + \lambda_{n+3} > 5(\sqrt{5} + 1)/4.$$

Hence

$$\lambda_n + \lambda_{n+1} + \lambda_{n+2} + \lambda_{n+3} > 5(\sqrt{5} + 1)/2 > (21 + 5\sqrt{5})/4. \tag{43}$$

(b) $\min(\lambda_n, \lambda_{n+1}) > \sqrt{5}$.

Here we have

$$\lambda_n + \lambda_{n+1} > 2\sqrt{5},$$
$$\lambda_{n+2} + \lambda_{n+3} > 4.$$

Hence

$$\lambda_n + \lambda_{n+1} + \lambda_{n+2} + \lambda_{n+3} > 4 + 2\sqrt{5} > 5(\sqrt{5} + 1)/2 > (21 + 5\sqrt{5})/4. \tag{44}$$

(c) $\min(\lambda_{n+2}, \lambda_{n+3}) > \sqrt{5}$.

The proof here is exactly as in case (b), and so the proof is complete.
From Theorem 13 we obtain easily

THEOREM 14. *For an arbitrary irrational number ξ, we have*

$$\underline{\lim} \left(\sum_{i=0}^{m-1} \lambda_i\right)/m \geqq (21 + 5\sqrt{5})/16 > 2.$$

Proof. Let k be an integer such that $4k \leqq m \leqq 4k+3$. Then

$$\Big(\sum_{i=0}^{m-1}\lambda_i\Big)/m \geqq \Big(\sum_{i=0}^{4k-1}\lambda_i\Big)/m \geqq k(21+5\sqrt{5})/4m \geqq k(21+5\sqrt{5})/(16k+12).$$

Theorem 13 can be improved by using larger sequences of consecutive λ_i. We shall consider only sequences of eight.

THEOREM 15. *The sum of any eight consecutive* λ_i *is greater than* $(31+15\sqrt{5})/4$.

Proof. Let $\lambda_n, \lambda_{n+1}, \cdots, \lambda_{n+7}$ be the sequence in question. Since

$$(31+15\sqrt{5})/4 = (21+5\sqrt{5})/4 + 5(\sqrt{5}+1)/2,$$

it follows from (43) and (44) that we need to consider only the case in which the first four and the last four of the λ_i each contain only one term exceeding $\sqrt{5}$. In this case it follows from Borel's Theorem that $\lambda_{n+2} > \sqrt{5}$ and $\lambda_{n+5} > \sqrt{5}$, and all the others are less than $\sqrt{5}$. Thus, by Theorem 9, $\lambda_{n+2} > 3$ and $\lambda_{n+5} > 3$. Since $\lambda_n + \lambda_{n+1} > 4$, $\lambda_{n+3} + \lambda_{n+4} > 4$, and $\lambda_{n+6} + \lambda_{n+7} > 4$, we obtain

$$\sum_{i=0}^{7}\lambda_{n+i} > 18 > (31+15\sqrt{5})/4.$$

THEOREM 16. *We have*

$$\underline{\lim}\Big(\sum_{i=0}^{m-1}\lambda_i\Big)/m \geqq (31+15\sqrt{5})/32 > 2.0169.$$

It follows from Hurwitz' Theorem that this lower limit cannot exceed $\sqrt{5}$ for certain irrational numbers ξ.

UNIVERSITY OF NORTH CAROLINA.

BIBLIOGRAPHY.

[1] E. Borel, "Contribution à l'analyse arithmétique du continu," *Journal de Mathématiques pures et appliquées*, ser. 5, vol. 9 (1903), pp. 329-375.

[2] L. R. Ford, "A geometrical proof of a theorem of Hurwitz," *Proceedings of the Edinburgh Mathematical Society*, vol. 35 (1916), pp. 59-65.

[3] M. Fujiwara, "Bemerkung zur Theorie der Approximation der irrationalen Zahlen durch rationale Zahlen," *Science Reports of the Tôhoku Imperial University*, ser. 1, vol. 13 (1924), pp. 1-11.

[4] ———, "Remarks on the theory of approximation of irrational numbers by

rational numbers," *Japanese Journal of Mathematics*, vol. 1 (1924), pp. 15-16.

[5] S. Fukasawa (Morimoto), "über die Kleinsche geometrische Darstellung des Kettensbruchs," *Japanese Journal of Mathematics*, vol. 2 (1926), pp. 101-114.

[6] ———, "Beweise einiger Sätze in der Kettenbruchstheorie durch die Humbertsche geometrische Darstellung," *Japanese Journal of Mathematics*, vol. 7 (1930), pp. 305-314.

[7] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford, Oxford University Press, 2nd edition, 1945.

[8] P. J. Heawood, "The classification of rational approximations," *Proceedings of the London Mathematical Society*, (2), vol. 20 (1922), pp. 233-250.

[9] G. Humbert, "Remarques sur certaines suites d'approximation," *Journal de Mathématiques pures et appliquées*, ser. 7, vol. 2 (1916), pp. 155-167.

[10] A. Hurwitz, "Über die angenäherte Darstellung, der Irrationalzahlen durch rationale Brüche," *Mathematische Annalen*, vol. 39 (1891), pp. 279-284.

[11] J. F. Koksma, *Diophantische Approximationen*, Berlin, Julius Springer, 1936.

[12] K. Kurosu, "Note on the theory of approximation of irrational numbers by rational numbers," *Tôhoku Mathematical Journal*, vol. 21 (1922), pp. 247-260.

[13] O. Perron, *Die Lehre von den Kettenbrüchen*, Leipzig, Teubner, 2nd edition, 1929.

[14] ———, "Über die Approximation irrationaler Zahlen durch rationale," I, II, *Sitzungsberichte der Heidelberger Akademie der Wissenschaften, Mathematisch-naturwissenschaftliche Klasse, Abteilung A*, 1921, no. 4 and no. 8.

[15] K. Shibata, "On the approximation of irrational numbers by rational numbers," *Tôhoku Mathematical Journal*, vol. 23 (1924), pp. 328-337.

[16] K. T. Vahlen, "Über Näherungswerte und Kettenbrüche," *Journal für die reine und angewandte Mathematik*, vol. 115 (1895), pp. 221-233.

ON LINEAR REPULSIVE FORCES.*

By Aurel Wintner.

1. The following theorem, and a refinement of it, will be proved:

In a linear, non-conservative system, of reversible type and of n degrees of freedom, let the kinetic energy be determined by n constant, positive mass factors, and let the potential energy be a non-positive definite quadratic form at every fixed t, with a coefficient matrix which varies with t continuously; $0 \leqq t < \infty$. *Then, no matter what the behavior, as* $t \to \infty$, *of this matrix may be, there exist solutions which are bounded as* $t \to \infty$. *More specifically, n (out of a total of 2n) linearly independent solutions will be of this type.*

The situation is illustrated in the trivial conservative case, in which the general solution is of the form

$$\sum_{k=1}^{n} (a_k e^{-\lambda_k t} + b_k e^{\lambda_k t}),$$

where a_k, b_k are $2n$-ary constant vectors, and $\lambda_1, \cdots, \lambda_n$ denote non-negative exponents which are independent of t and of the integration constants. Another illustration is the non-conservative case with a single degree of freedom. Then the theorem reduces to a well-known result of A. Kneser[1] concerning the scalar differential equation $x'' = p(t)x$, where $p(t) > 0$ is continuous for $0 \leqq t < \infty$.

2. Since the kinetic energy is of the form $\frac{1}{2} \Sigma\, m_k x_k'^2$, where every m_k is positive and independent of t, there is no loss of generality in assuming that every m_k is 1. In fact, the substitution $\Sigma\, m_k x_k'^2 \to \Sigma\, x_k'^2$ can be effected by a linear transformation of the coordinates $x_1, \cdots, x_n$ which is independent of t and has a non-vanishing determinant; so that, by the inertia theorem, the signature of the quadratic form representing the potential energy remains unaltered at every t.

In the resulting normalization, the Lagrangian equations can be written in the form

$$x'' - F(t)x = 0, \tag{1}$$

---

* Received July 19, 1948.

[1] A. Kneser, "Untersuchung und asymptotische Darstellung der Integrale gewisser Differentialgleichungen bei grossen reellen Werthen des Arguments, I," *Journal für die reine und angewandte Mathematik*, vol. 116 (1896), pp. 178-212.

where $F(t)$ is a real, symmetric, non-negative definite, n-rowed matrix which is a continuous function of t, and x is the coordinate vector $(x_1, \cdots, x_n)$.

If a dot denotes scalar multiplication of vectors, the quadratic form belonging to the matrix $F(t)$ is $x \cdot F(t)x$. Since this form is supposed to be incapable of negative values, it follows from (1) that the inequality $x \cdot x'' \geqq 0$ holds for every t, if $x(t)$ is any solution vector (with real components). On the other hand, $x \cdot x'' = (\frac{1}{2}x^2)'' - x'^2$, where $y^2 = y \cdot y = |\, y\, |^2$ for $y = x, x'$. It follows therefore from $x \cdot x'' \geqq 0$ that

$$r'' \geqq 0, \text{ where } r = |\, x\, |^2. \tag{2}$$

3. The last inequality means that the squared length of any solution vector $x(t)$, when plotted against the t-axis, is everywhere convex toward the t-axis. Since the curve $r = r(t)$ must be above or on the t-axis, it follows that, if the trivial solution, $x(t) \equiv 0$, of (1) is excluded, the curve $r = r(t)$ cannot have more than one of its points on the t-axis. Accordingly, if $0 \leqq \alpha < \beta < \infty$, a solution vector $x(t)$ satisfies both n-ary conditions

$$x(\alpha) = 0, \qquad x(\beta) = 0 \tag{3}$$

only if $x(t) \equiv 0$.

It follows that, if a, b is any pair of points in the x-space, there exists a unique solution vector $x = x(t) = x_{ab}(t)$ satisfying the boundary conditions

$$x(\alpha) = a, \qquad x(\beta) = b. \tag{4}$$

In fact, if $x^1(t), \cdots, x^{2n}(t)$ is any fixed system of linearly independent solutions of (1), every solution $x(t)$ determines a unique set of $2n$ scalar integration constants c_i satisfying

$$x(t) = \sum_{i=1}^{2n} c_i x^i(t)\,; \tag{5}$$

conversely, (5) is a solution for every choice of $c_1, \cdots, c_{2n}$. Hence, if a and b are given, (4) represents $2n$ linear equations for $2n$ unknowns c_i, and so the assertion is equivalent to the statement that the determinant of these equations is not 0. But if this determinant were 0, the homogeneous equations, those represented by the case (3) of (4), had a solution $(c_1, \cdots, c_{2n}) \neq (0, \cdots, 0)$. Since such a set of integration constants c_i defines a solution (5) distinct from $x(t) \equiv 0$, there results a contradiction to the fact that (3) holds only if $x(t) \equiv 0$.

Clearly, this argument is just the n-dimensional formulation of Jacobi's proof for the non-existence of conjugate points on the geodesics of surfaces of non-positive Gaussian curvature.

4. Choose in (4)

$$\alpha = 0, \qquad \beta = m; \qquad\qquad a = e, \qquad b = 0, \tag{6}$$

where m is a positive integer and e a point of the unit sphere $|x| = 1$. For an arbirarily fixed e, let ${}^m x(t)$, where $0 \leqq t < \infty$, denote the solution $x(t)$ determined by (6) and (4). Then, as shown above, none of the functions

$$r_m = r_m(t) = |\,{}^m x(t)\,|^2 \tag{7}$$

can have two zeros. On the other hand, each of the functions (7) has a zero; in fact, by (6) and (4),

$$r_m(0) = 1, \qquad r_m(m) = 0. \tag{8}$$

But (8) and the convexity condition (2) imply that $r_m(t)$ is monotone for $0 \leqq t < m$. Accordingly,

$$r_m(t) > 0, \qquad r'_m(t) \leqq 0, \qquad r''_m(t) \geqq 0 \text{ if } 0 \leqq t < m. \tag{9}$$

The first of the conditions (8) and the second of the inequalities (9) imply that the functions $r_m(t)$ are uniformly bounded on every fixed, bounded t-interval. In view of (7), the same holds for the vectors ${}^m x(t)$ and so, by (1), for the vectors ${}^m x''(t)$ as well, the elements of the matrix $F(t)$ being continuous, hence locally bounded. On the other hand, the uniform boundedness of ${}^m x''(t)$ for $0 < t < 1$ implies, by Taylor's theorem, the boundedness of ${}^m x(1) - {}^m x(0) - {}^m x'(0)$ as $m \rightarrow \infty$. Since both ${}^m x(1)$ and ${}^m x(0)$ are bounded, this means that ${}^m x'(0)$ is bounded, as $m \rightarrow \infty$. Hence, a quadrature shows that the uniform boundedness of ${}^m x''(t)$ on every fixed, bounded t-interval implies the same for ${}^m x'(t)$.

Accordingly, both sequences ${}^1 x(t), {}^2 x(t), \cdots, {}^1 x'(t), {}^2 x'(t), \cdots$ are uniformly bounded, and have uniformly bounded derivatives, and are therefore equicontinuous, on every fixed, bounded t-interval. Consequently, the sequence $1, 2, \cdots$ contains a subsequence having the property that, if the m-th element of the subsequence is denoted simply by m, then ${}^m x(t)$ and ${}^m x'(t)$ tend, as $m \rightarrow \infty$, to certain functions, $x(t)$ and $x'(t)$, uniformly on every fixed, bounded interval contained in the half-line $0 \leqq t < \infty$, and the limit $x'(t)$ is the derivative of the limit $x(t)$.

5. Since every ${}^m x(t)$ is a solution of (1), it is clear that the limit function $x(t)$, just obtained, is a solution of (1), and that, corresponding to (9) and (7),

$$(10) \qquad r(t) > 0, \qquad r'(t) \leqq 0, \qquad r''(t) \geqq 0 \text{ for } 0 \leqq t < \infty,$$

where $r(t) = | x(t) |^2$. Hence, the theorem italicized above can be refined as follows:

If $F(t)$, *where* $0 \leqq t < \infty$, *is an n-rowed, real, symmetric matrix of continuous functions which is non-negative definite at every fixed t, then the system* (1) *has n linearly independent solution vectors* $x(t)$ *which are "asymptotic solutions," in the sense that* $r(t) = | x(t) |^2$ *satisfies* (10); *so that, in particular,*

$$\lim_{t\to\infty} r(t) \text{ exists } (< \infty) \text{ and } \lim_{t\to\infty} r'(t) = 0.$$

In fact, since every ${}^m x(0)$ was chosen to be e, an arbitrary vector of length 1, and since ${}^m x(t) \to x(t)$ as $m \to \infty$ holds for every t and so for $t = 0$, the initial vector $x(0)$ of the solution occurring in (10) can be chosen to be any vector e of length 1. Hence, there are admitted n linearly independent initial vectors $x(0)$. But then, no matter what the initial vectors $x'(0)$ may be, there must result n linearly independent solution vectors $x(t)$.

6. If A is any matrix, let A^* denote the transposed matrix.

The preceding theorem remains true if $F(t)$ *in* (1) *is replaced by* $F(t) + S(t)$, *where* $S(t)$ *is any matrix which is skew-symmetric and continuous at every t.*

In particular, instead of the assumption that $F(t)$ *in* (1) *is symmetric and non-negative definite, it is sufficient to assume that* $F(t)$ *in* (1) *is such as to make the matrix* $F(t) + F^*(t)$ *non-negative definite.*

The second of these assertions follows from the first, since $F + S$ becomes $\frac{1}{2}(F + F^*)$ if $S = \frac{1}{2}(F^* - F)$.

In order to verify the first assertion, let (1) be replaced by

$$x'' - (F + S)x = 0.$$

Then, since $x \cdot Sx = 0$ is an identity for every skew-symmetric S,

$$x \cdot x'' = x \cdot Fx.$$

This leads to (2), as above, and the balance of the proof remains unchanged.

Interesting is the particular case $x'' = S(t)x$, where $S(t)$ is any real, skew-symmetric matrix of continuous functions. This case is covered by the above results, since the zero matrix, being non-negative definite, can be chosen to be $F(t)$ in $F(t)+S(t)$.

7. The result concerning the number of "asymptotic" (or, for that matter, just bounded) linearly independent solutions cannot be improved. In fact, under any of the above assumptions, *there exist* n *(out of* $2n$*) linearly independent solutions* $x(t)$ *satisfying* $|x(t)| \to \infty$ *as* $t \to \infty$.

First, it is clear that there exist n linearly independent $2n$-ary vectors $v_1, \cdots, v_n$ having the following property: If $v = (c_1, \cdots, c_{2n})$ is any of the $2n$-ary vectors $v_1, \cdots, v_n$, then $a \cdot b > 0$ holds for the pair of n-ary vectors defined by $a = (c_1, \cdots, c_n)$, $b = (c_{n+1}, \cdots, c_{2n})$.

Corresponding to any such a, b, let $x(t)$ denote the solution defined by the initial conditions $x(0) = a$, $x'(0) = b$. Then $r'(0) = 2a \cdot b$, since $r(t) = x(t) \cdot x(t)$. Consequently, $r'(t) > 0$ holds at $t = 0$ and therefore at every t close enough to $t = 0$. It follows therefore from (2) that, for reasons of convexity, $r(t) \to \infty$ as $t \to \infty$.

THE JOHNS HOPKINS UNIVERSITY.

ON THE LAPLACE-FOURIER TRANSCENDENTS.*

By PHILIP HARTMAN and AUREL WINTNER.

The present paper deals with extensions of the results proved on pp. 87-96 of vol. 69 (1947) of this Journal. The nature of the extensions in question can be illustrated by the following theorem, proved *loc. cit.* only for the particular case $a_0 = a_1 = \cdots = 0$:

If a_n, b_n are real, non-negative numbers corresponding to which the coefficient functions of the differential equation

$$(1) \qquad x'' + x' \sum_{n=0}^{\infty} (-1)^n a_n/t^{2n+1} + x \sum_{n=0}^{\infty} (-1)^n b_n/t^{2n} = 0$$

are entire functions in $1/t$, then (1) has on the half-line

$$(2) \qquad 0 < t < \infty$$

a solution of the form

$$(3) \qquad x(t) = (\int_0^{\infty}) e^{-itu} d\alpha(u) \not\equiv 0,$$

where $\alpha(u)$ is monotone (but not necessarily bounded) for $0 \leqq u < \infty$, and the convergence of (3) on (2) is part of the statement.

In (3), the parenthesis of the integral sign refers to the Abelian evaluation of

$$\int_0^{\infty} e^{-itu} d\alpha(u), \quad \text{as} \quad \lim_{\epsilon \to +0} \int_0^{\infty} e^{-\epsilon u} e^{-itu} d\alpha(u).$$

If k is a real index, not necessarily an integer, then Bessel's equation,

$$t^2 x'' + t x' + (t^2 - k^2) x = 0,$$

supplies an illustration of the theorem, since the coefficients a_n, b_n in (1) become

$$a_0 = 1, \quad a_1 = a_2 = \cdots 0, \quad b_0 = 1, \quad b_1 = k^2, \quad b_2 = b_3 = \cdots = 0;$$

cf. *loc. cit.*, p. 89 (the footnote on p. 89 is relevant in the present case also).

If t is replaced by ti, then (1) goes over into

$$(4) \qquad x'' + r(t) x' - q(t) x = 0,$$

---

* Received September 1, 1948.

where

$$r(t) = \sum_{n=0}^{\infty} a_n/t^{2n+1}, \qquad q(t) = \sum_{n=0}^{\infty} b_n/t^{2n}. \tag{5}$$

If $m = 0, 1, 2, \cdots$, then $(-1)^m$ times the m-th derivative of either of the functions (5) is non-negative on the half-line (2), since $a_n \geqq 0$, $b_n \geqq 0$. In other words, both functions (5) are completely monotone on (2). It will be shown that this alone implies for (4) the existence of a solution of the form

$$x(t) = \int_0^{\infty} e^{-tu} d\alpha(u) \neq 0 \tag{6}$$

on (2), where $\alpha(u)$ is monotone for $0 \leqq u < \infty$. For reasons of analyticity, (6) must be a solution of (4) when t, instead of being positive, is complex with a positive real part. In particular, (6) will represent a solution of (4) on the line $\epsilon + it$, where $\epsilon > 0$ is fixed and t is a real variable. Hence, the existence of a solution (3) of (1) follows by letting $\epsilon \rightarrow 0$.

The proof for the existence of a solution (6) of (4) will be based on the following

LEMMA. *On an interval*

$$0 \leqq t < T, \tag{7}$$

where the case of a half-line ($T = \infty$) *is allowed, let* $p(t)$, $r(t)$, $q(t)$ *be real-valued, continuous functions, the first of which is positive, the second arbitrary, the third non-negative. Then the differential equation*

$$p(t)x'' + r(t)x' - q(t)x = 0 \tag{8}$$

has a solution which is positive and non-increasing on (7).

The same is true if (8) *is replaced by*

$$(P(t)x')' - Q(t)x = 0, \tag{9}$$

where $P(t) > 0$ *and* $Q(t) \geqq 0$ *are continuous on* (7).

REMARK. *The assertions of the Lemma remain true if* (7) *is replaced by the open interval* $0 < t < T$ (*where, as before,* $T = \infty$ *is allowed*).

It will be noted that (9) becomes a particular case,

$$p = P, \qquad r = P', \qquad q = Q,$$

of (8) only if P, instead of being just continuous, is required to have a continuous derivative.

If (8) reduces to

$$x'' - q(t)x = 0, \text{ where } q(t) \geqq 0, \tag{10}$$

the assertion of the Lemma becomes identical with the result of A. Kneser (1896), used *loc. cit.*

In the case of (9), replace t by $s = s(t)$, where

$$s = \int_0^t dv/P(v).$$

Then s is an increasing function of t, and (9) appears in the form (10), if the independent variable of (10) is s, and q denotes the product PQ (as a function of s). Since $PQ \geqq 0$, the assumption, $q \geqq 0$, of (10) is satisfied.

In the case of (8), put

$$P(t) = \exp \int_0^t r(v)\,dv/p(v)$$

and multiply (8) by $P(t)/p(t)$ (which is positive throughout). Then (8) appears in the form (9), where $Q(t) = q(t)P(t)/p(t)$ (hence $Q(t) > 0$).

This proves the Lemma. The Remark following the Lemma results if (7) is replaced by the interval $\epsilon \leqq t < T$, where $\epsilon > 0$, and, after the Lemma has been applied on the latter interval, ϵ tends to 0.

By the procedure applied *loc. cit.*, there will now be proved the main

THEOREM. *Let two functions, $q(t)$ and $r(t)$, and the first derivative, $p'(t)$, of a third positive function, $p(t)$, be completely monotone on the half-line* (2). *Then the differential equation* (8) *has a solution $x(t) \neq 0$ which is completely monotone on* (2), i. e., which is representable on (2) in the form (6) (Hausdorff-Bernstein), where $\alpha(u)$ is non-decreasing, but not necessarily bounded, for $0 \leqq u < \infty$.

COROLLARY. *If $p(t)$ is positive on* (2) *and if $q(t)$ and the first derivative of $p(t)$ are completely monotone on* (2), *then the self-adjoint differential equation*

$$(p(t)x')' - q(t)x = 0$$

has a solution $x(t) \neq 0$ which is completely monotone on (2).

In fact, the assumptions of the Corollary imply those of the Theorem, $r(t)$ being $p'(t)$.

In the particular case (10) of (8), the assertion of the Theorem was proved *loc. cit.* Although (8) is a normal form of (10), the reduction of the assertion of the Theorem to this normal form leads to difficulties. For this reason, the following proof for the general case will proceed directly.

First, since $p > 0$ and $q \geqq 0$ are continuous on (2), the Remark following the Lemma assures for (8) the existence of a solution $x(t) \neq 0$ satisfying

$$(11) \qquad (-1)^n d^n x/dt^n \geqq 0 \text{ for } n = 0 \text{ and } n = 1, \text{ where } 0 < t < \infty.$$

Next, if $x = -t$, then (2) goes over into $-\infty < s < 0$, and (8) into

$$(12) \qquad P(s)\, d^2x/ds^2 = R(s)\, dx/ds + Q(s)x,$$

where

$$(13) \qquad P(s) = p(-s), \qquad R(s) = r(-s), \qquad Q(s) = q(s).$$

Since $p(t)$ is positive, and $dp(t)/dt$, $r(t)$, $q(t)$ are completely monotone, on (2), it is clear from (13) that $P(s)$ is positive, that each but the 0-th derivative of $P(s)$ is non-positive, and that each (including the 0-th) derivative of $R(s)$ and $Q(s)$ is non-negative (if the derivatives are taken with respect to s, where $-\infty < s < 0$). Hence, if (12) is differentiated n times with respect to s, it is seen that the product $Pd^{n+2}x/ds^{n+2}$ is a linear form, with non-negative coefficients, in

$$(14) \qquad x,\ dx/ds,\ d^2x/ds^2, \cdots, d^nx/ds^n,\ d^{n+1}x/ds^{n+1}.$$

Since $P > 0$, it follows that $d^{n+2}x/ds^{n+2}$ is non-negative if each of the $n + 2$ functions (14) is non-negative. Consequently,

$$(15) \qquad d^nx/ds^n \geqq 0, \text{ where } -\infty < s < 0,$$

holds for every n, if (15) holds for $n = 0$ and for $n = 1$.

Since $t = -s$, the assertion of (11) is that (15) holds for $n = 0$ and for $n = 1$. Hence, (15) holds for every $n > 0$. In view of $t = -s$, this means that

$$(16) \qquad (-1)^n d^nx/dt^n \geqq 0, \text{ where } 0 < t < \infty$$

holds for every $n > 0$.

According to (16), the solution $x(t)$ of (8) is completely monotone on (2). Hence, in order to complete the proof of the Theorem, it is sufficient to ascertain that $x(t) > 0$. But this follows from the choice of $x(t)$, according to which $x(t)$ is positive (at every $t > 0$).

APPENDIX.

As emphasized above, the non-decreasing function $\alpha(u) \neq$ const. occurring in (6) is not in general bounded as $u \rightarrow \infty$. A by-product of the following considerations will be a characterization of the coefficient functions $p(t)$, $r(t)$, $q(t)$ of those differential equations (8) corresponding to which the function $\alpha(u)$ occurring in the solution (6) is bounded as $u \rightarrow \infty$.

First, since $\alpha(u)$ is non-decreasing, it is clear from (6) that $\alpha(u)$ is bounded as $u \rightarrow \infty$ if and only if $x(t)$ is bounded as $t \rightarrow +0$. But the boundedness of $x(t)$ near $t = 0$, being an issue of purely local nature, has nothing to do with the question of complete monotony. Hence, it is sufficient to answer the question for the case in which (8) is given in its normal form (10). The answer then runs as follows:

Let $q(t)$ *be real-valued, continuous and non-negative for* $0 < t \leqq 1$. *Then, according to the Lemma,*

$$x'' - q(t)x = 0 \tag{17}$$

has a solution satisfying

$$x(t) > 0 \text{ and } x'(t) \leqq 0 \text{ for } 0 < t \leqq 1. \tag{18}$$

In order that this solution be bounded (*as* $t \rightarrow +0$), *the condition*

$$\int_{+0}^{1} tq(t)\,dt < \infty \tag{19}$$

is necessary and sufficient.

The sufficiency of (19) is quite on the surface, in the sense that it has nothing to do with the assumption $q(t) \geqq 0$. In fact, if $q(t)$, where $0 < t \leqq 1$, is *any* continuous function, then

$$\int_{+0}^{1} t\,|\,q(t)\,|\,dt < \infty \tag{19 bis}$$

is sufficient in order that every solution $x(t)$ of (17) be such as to have a finite limit $x(+0)$. This is contained in Bôcher's extension of the theory of Fuchsian points to the real field (Trans. Amer. Math. Soc., vol. 1 (1900), pp. 40-52, § 4).

The proof of the converse depends on the consideration of the Lagrangian function, say $\frac{1}{2}L$, of (17) along an arbitrary solution $x(t)$ of (17); so that $L = L(t)$ and

$$L(t) = x'^2(t) + q(t)x^2(t). \tag{20}$$

Even if $q \geqq 0$ is not assumed,

$$tx(t)x'(t) = \text{const.} + \tfrac{1}{2}x^2(t) - \int_t^1 sL(s)\,ds \tag{21}$$

holds for $0 < t \leqq 1$. In fact, it is easily verified by differentiation that (21) is an identity in t by virtue of (17) and (20).

It will be concluded from (21) that, if $x(t)$ is a bounded solution satisfying (18) and if $q \geqq 0$, then *the improper integral*

$$\int_{+0}^{1} tL(t)\,dt \text{ is convergent.} \tag{22}$$

In view of (20), where $q \geqq 0$, this will imply that

$$\int_{+0}^{1} tq(t)x^2(t)\,dt < \infty.$$

Hence, (19) will follow from the fact that, according to (18), the function $x(t)$ must keep away from 0 as $t \to +0$.

Actually, (18) and the boundedness of $x(t)$ assure the existence of a positive limit $x(+0) \neq \infty$. Hence,

$$x^2(1) - x^2(+0) = \int_{+0}^{1} dx^2(t) = 2\int_{+0}^{1} x(t)x'(t)\,dt.$$

Suppose that $x(t)x'(t)$ is monotone. Then the convergence of the last integral implies that

$$tx(t)x'(t) \to 0 \text{ as } t \to +0.$$

This limit relation, when combined with (21), where $x^2(+0)$ is supposed to exist ($\neq \infty$), contains the truth of (22). Hence, the proof will be complete if it is ascertained that $x(t)x'(t)$ is monotone.

To this end, the assumption $q \geqq 0$ will be used again. It implies, by (17), that $xx'' \geqq 0$. Since $(xx')' = xx'' + x'^2$, it follows that $(xx')' \geqq 0$. Hence, $x(t)x'(t)$ is non-decreasing.

THE JOHNS HOPKINS UNIVERSITY.

FURTHER CONGRUENCE PROPERTIES OF THE FOURIER COEFFICIENTS OF THE MODULAR INVARIANT $j(\tau)$.*

By JOSEPH LEHNER.

This article is a continuation of a previous one [2], in which it was shown that the Fourier coefficients, c_n, of the modular invariant $j(\tau)$,

$$j(\tau) = x^{-1} + 744 + 196,884x + \cdots, \qquad x = \exp 2\pi i\tau,\ I(\tau) > 0,$$

enjoy certain congruence properties with respect to the moduli 5, 7 and their powers, and 11, 11^2. In this paper we shall consider the moduli 2^α, 3^α, $\alpha \geq 1$, and prove

THEOREM 1. *If*

$$j(\tau) = x^{-1} + c_0 + \sum_{n=1}^{\infty} c_n x^n, \qquad x = \exp 2\pi i\tau \tag{1}$$

then

$$c_n \equiv 0 \pmod{2^{3\alpha+8}} \text{ if } n \equiv 0 \pmod{2^\alpha} \tag{2}$$

$$c_n \equiv 0 \pmod{3^{2\alpha+3}} \text{ if } n \equiv 0 \pmod{3^\alpha} \tag{3}$$

for $\alpha, n = 1, 2, 3, \cdots$.

The method of proof is that used in [2]. The congruences follow from *identities*, in which the left members are series

$$\sum_{\mu=0}^{\infty} c_{n\mu} x^\mu, \qquad n \equiv 0 \pmod{p^\alpha},\ p = 2 \text{ or } 3$$

while the right members are polynomials, with rational integral coefficients, in certain functions Φ, which themselves have series expansions with integral coefficients.

The proof is in two parts. The first part consists in showing that for $\alpha = 1$ (in Theorem 1), both members of the identity described above are modular functions belonging to a certain subgroup of the modular group. The identity is then established by matching the principal parts of the two members at the parabolic vertices of the subgroup. 1 is concerned with this.

The second part of the proof proceeds by induction on α. Here the *modular equation* for the function Φ plays an essential role. The modular equation is an algebraic equation connecting $\Phi(\tau)$ and $\Phi(p\tau)$. For $p = 5, 7$,

---

* Received September 3, 1948.

these equations were worked out by Rademacher ([3], pp. 626, 628). It seems worth while to give here a systematic method which is applicable to all primes of interest, viz., $p = 2, 3, 5, 7, 13$. This is done in **2** and the modular equations given. As will appear in the course of the argument, the above primes are those for which the subgroup $\Gamma_0(p)$ is of genus zero. ($\Gamma_0(p)$ is defined by the condition $p \mid c$ in the modular substitution $\tau' = (a\tau + b)/(c\tau + d)$.) The induction on α is carried out in **3** and completes the proof of Theorem 1.

In **5** we consider modular functions other than $j(\tau)$ which share with it the congruence properties of Theorem 1 and of Theorems 1 and 2 of [2].

1. Congruences for the moduli 2 and 3. We are going to follow closely the reasoning in [2]. We recall that $F(\tau)$ is a modular function on a group Γ if

1) it is a meromorphic function of τ in the half-plane $I(\tau) > 0$;
2) it possesses only poles in the local variable at the parabolic vertices of Γ;
3) it is invariant: $F(V\tau) = F(\tau)$, for every modular substitution $V\tau = (a\tau + b)/(c\tau + d)$ of Γ.

In what follows, Γ may be either the modular group or one of its subgroups of finite index. $\Gamma_0(p)$ is the subgroup given by $c \equiv 0 \pmod{p}$. p is one of the primes 2, 3, 5, 7, 13.

We introduce the linear operator

$$U_p f(\tau) = p^{-1} \sum_{\lambda=0}^{p-1} f((\tau + \lambda)/p) \tag{1.1}$$

where $f(\tau)$ is given by the expansion

$$f(\tau) = \sum_{n=-s}^{\infty} a_n x^n,$$

and throughout this paper

$$x = \exp 2\pi i\tau, \qquad I(\tau) > 0. \tag{1.2}$$

In [2], **2** it is shown that

$$U_p f(\tau) = \sum_{n=t}^{\infty} a_{pn} x^n, \qquad t = -[s/p]. \tag{1.3}$$

Moreover, it is proved that if $f(\tau)$ is a modular function belonging to the full modular group, then U_pf belongs to the subgroup $\Gamma_0(p)$. (The proof was carried out for the function $j(\tau)$ but is easily seen to be valid for any $f(\tau)$ on the full group.)

A formula which will be useful later is (8.81) of [1]:

$$(1.4) \qquad pU_pF(-1/p\tau) - pU_pF(p\tau) = F(-1/p^2\tau) - F(\tau),$$

where $F(\tau)$ is an entire modular function of $\Gamma_0(p)$. By $U_pF(p\tau)$ we mean $U_pF(\tau)$ with τ replaced by $p\tau$; similarly for $U_pF(-1/p\tau)$.

Since the subgroups $\Gamma_0(p)$ are of genus zero for the primes of interest, they will possess univalent functions, which may be taken as

$$(1.5) \qquad \Phi_{p,r}(\tau) = \Phi(\tau) = \{\eta(p\tau)/\eta(\tau)\}^r = x + \cdots,$$

with

$$(1.6) \qquad \eta(\tau) = \exp(\pi i\tau/12) \prod_1^\infty (1 - x^m),$$

and

$$(1.7) \qquad r(p-1) = 24.$$

That the functions Φ belong to $\Gamma_0(p)$ has been demonstrated by Rademacher for $p > 3$ ([3], Theorem 1, p. 619) if we keep in mind that, for the primes under consideration in this paper, r is always even. But his proofs, with minor modifications, are readily seen to hold for[1] $p = 2, 3$.

The functions Φ are clearly univalent in the fundamental region of $\Gamma_0(p)$, as is evident from the following facts: 1) the fundamental region of $\Gamma_0(p)$ has two parabolic points, say $\tau = i\infty, 0$; 2) Φ has a simple zero at $\tau = i\infty$ and a simple pole at $\tau = 0$ measured in $x = \exp 2\pi i\tau$; 3) Φ is regular and zero free in the interior of the fundamental region. These considerations are developed in detail in [2] for $p > 3$ (2) but are equally valid for $p \leq 3$.

With these preliminaries out of the way, we are ready to express $U_pj(\tau)$ as a polynomial in $\Phi(\tau)$. For this purpose we need consider the vertex at $\tau = 0$ only, since U_pj and Φ are both regular at $\tau = i\infty$ ([2], (2.3)). From [2], (2.4), (2.6) we take the expansions, valid for $p \leq 3$,

$$(1.8) \qquad pU_pj(\tau') = x^{-p^2} + \cdots$$

$$(1.9) \qquad p^{r/2}\Phi(\tau') = x^{-1} + \cdots, \qquad \tau' = -1/p\tau,$$

---

[1] When $p = 2$, $r = 24$, and $\Phi = \{\eta(2\tau)/\eta(\tau)\}^{24} = \Delta(2\tau)/\Delta(\tau)$, where Δ is the "discriminant" of elliptic function theory. Δ is a modular *form* of dimension -12; its transformation equation is $\Delta(V\tau) = (c\tau + d)^{12}\Delta(\tau)$. It follows easily that $\Delta(2\tau)/\Delta(\tau)$ is invariant on $\Gamma_0(2)$.

the coefficients in the right members being integers. From a comparison of the principal parts it follows that U_pj is equal to a polynomial in Φ whose coefficients, except possibly for the constant term, are integers divisible by $p^{r/2-1}$:

$$U_pj(\tau)=C_0+p^{r/2-1}\sum_{h=1}^{p^2}C_hp^{r(h-1)/2}\Phi^h(\tau), \tag{1.10}$$

with integral $C_1, C_2, \cdots$.

Since for $p=2$, $r=24$; for $p=3$, $r=12$; Theorem 1 for $\alpha=1$ follows at once from (1.10) and (1.4).

2. **Modular equations.**

In order to generalize the congruences just proved to powers of 2 and 3 as moduli, it will be necessary to introduce the modular equation of Φ. The required results are contained in

THEOREM 2. *Let the genus of* $\Gamma_0(p)$ *be zero. Put*

$$Z=\Phi_{p,r}(\tau), \qquad W=p^{r/2}\Phi_{p,r}(\tau/p).$$

Then

$$U_pZ=p\sum_{j=1}^{p}b_jZ^j. \tag{2.1}$$

Moreover, the modular equation connecting Z *and* W *is*

$$W^p+\sum_{j=1}^{p}(-1)^jp_jW^{p-j}=0, \tag{2.2}$$

where

$$(-1)^{j+1}p_j=p^{r/2+2}\sum_{l=j}^{p}b_lZ^{l-j+1}. \tag{2.3}$$

Here, the b_j *are integers with one exception: for* $p=13$, $13b_1$ *is an integer.*

Proof. The proof follows a method sketched in a previous paper ([1], p. 511). Applying (1.4) with $F=\Phi_{p,r}(\tau)=Z$, we have

$$pU_pZ(-1/p\tau)-pU_pZ(p\tau)=Z(-1/p^2\tau)-Z(\tau). \tag{2.4}$$

Now assuming the truth of (2.1) for the moment, we get

$$\begin{aligned} pU_pZ(-1/p\tau)&=p^2\sum_{j=1}^{p}b_jZ(-1/p\tau)^j \\ &=p^2\sum_{1}^{p}b_jp^{-rj/2}Z(\tau)^{-j}, \end{aligned} \tag{2.5}$$

where we have used the property

$$Z(-1/p\tau) = p^{-r/2}Z^{-1}(\tau), \tag{2.6}$$

as given in [1], (8.83). Also

$$pU_pZ(p\tau) = p^2\sum_{1}^{p} b_jZ(p\tau)^j, \tag{2.7}$$

and

$$Z(-1/p^2\tau) = p^{-r/2}Z^{-1}(p\tau), \tag{2.8}$$

as follows from (2.1) and (2.6) on replacing τ by $p\tau$.

Substituting (2.5), (2.7), and (2.8) in (2.4) gives, after replacing τ by τ/p,

$$p^{r/2+2}\sum_{j=1}^{p} b_j(W^{-j} - Z^j) + W - Z^{-1} = 0, \tag{2.9}$$

an equation whose left member obviously has the factor $W^{-1} - Z$. If we exclude this factor, multiply by W^{p-1}, and rearrange, we obtain

$$W^p - p^{r/2+2}\sum_{j=1}^{p} b_j \sum_{k=0}^{j-1} W^{p-k-1}Z^{j-k} = 0, \tag{2.10}$$

which, with a change of summation variables, is exactly (2.2) and (2.3).

The remainder of the proof is taken up with establishing (2.1) and calculating the b_j. First we show that U_pZ is a function of $\Gamma_0(p)$. Its invariance under the substitutions $V\tau$ of this subgroup follows from [1], Lemma 1, p. 501:

$$\begin{aligned} U_pZ(V\tau) &= p^{-1}\sum_{\lambda=0}^{p-1} Z(T_\lambda V\tau) = p^{-1}\sum_{\mu=0}^{p-1} Z(W_\mu T_\mu \tau) \\ &= p^{-1}\sum_{\mu=0}^{p-1} Z(T_\mu\tau) = U_pZ(\tau), \qquad W_\mu \varepsilon \Gamma_0(p^2). \end{aligned} \tag{2.11}$$

Next, Z vanishes at $\tau = i\infty$; therefore, U_pZ does also. Finally, at $\tau' = 0$, we apply (2.4), (2.8) and have

$$\begin{aligned} pU_pZ(-1/p\tau) &= pU_pZ(p\tau) + p^{-r/2}Z^{-1}(p\tau) - Z(\tau) \\ &= p^{-r/2}x^{-p} + O(1), \end{aligned} \tag{2.12}$$

i. e., U_p has a pole of order p in x at $\tau' = 0$.

Since Z has a pole of first order at zero while U_pZ has a pole of order p there, it is clear from familiar arguments that U_pZ is a polynomial in Z of degree p. This polynomial will not have a constant term, since U_pZ and Z both vanish at infinity.

Write

$$Z(\tau)^j = x^j(1 + a_1^{(j)}x + a_2^{(j)}x^2 + \cdots), \tag{2.13}$$

and note that, by an elementary theorem on binomial coefficients,

$$p \mid a_1^{(p)}, a_2^{(p)}, \cdots, a_{p-1}^{(p)}. \tag{2.14}$$

Then, using (2.12) and (2.6), we have the following equations which secure agreement of the principal parts of the two members of (2.1) at $\tau' = 0$ and thus serve to determine the coefficients b_j:

$$\left\{\begin{array}{ll} p^{-1-r/2} & = b_p p^{1-rp/2} \\ 0 & = b_p p^{-rp/2} a_1^{(p)} + b_{p-1} p^{-r(p-1)/2} \\ \cdot & \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ 0 & = b_p p^{-rp/2} a_{p-1}^{(p)} + b_{p-1} p^{-r(p-1)/2} a_{p-2}^{(p-1)} + \cdots + b_1 p^{-r/2}. \end{array}\right. \tag{2.15}$$

By solving the system (2.15) recursively and remembering (2.14) one finds easily that

$$\begin{array}{ll} b_p = p^{r(p-1)/2-2} = p^{10}, & \\ p^{r(j-1)/2-1} \mid b_j, & j = 1, 2, \cdots, p-1. \end{array} \tag{2.16}$$

Hence, b_j is always an integer when $j > 1$, since $r \geq 2$; and pb_1 is an integer. Actually, it will turn out in the calculations that b_1 itself is integral except when $p = 13$.

It remains to compute the b_j. We list below for each value of p the expansion of $p^{r/2}Z(-1/p\tau)$ to p terms, from which the needed powers of Z are readily obtained. From these and (2.12), the b_j are calculated comparatively easily. The b_j are also listed.

The b_j shown in the table agree, for $p = 5, 7$, with the coefficients of Rademacher's modular equations ([3], pp. 626, 628). However, we have used a slightly different notation. If, in our equation (2.9), τ is replaced by 5τ, then the equation is multiplied by ZW^{-1}, it becomes Rademacher's equation following (11.4), with the identifications

$$X = W^{-1}(5\tau), \qquad Y = Z(5\tau), \qquad 5^2 b_j = \beta_j.$$

For $p = 7$, our formula (2.9) is directly comparable with Rademacher's equation (12.5).

VALUES OF b_j IN (2.1)

j \ p	2	3	5	7	13
1	3.2^2	10.3	63	82	$1{,}165.13^{-1}$
2	2^{10}	4.3^6	52.5^3	176.7^2	$9{,}604$
3		3^{10}	63.5^5	845.7^3	$27{,}272.13$
4			6.5^8	272.7^5	$41{,}140.13^2$
5			5^{10}	46.7^7	$3{,}014.13^4$
6				4.7^9	$25{,}660.13^4$
7				7^{10}	$12{,}086.13^5$
8					$4{,}180.13^6$
9					$1{,}064.13^7$
10					196.13^8
11					25.13^9
12					2.13^{10}
13					13^{10}

EXPANSION OF $p^{r/2}Z(-1/p\tau)$ IN POWERS OF $x = \exp 2\pi i\tau$

p \ Power of	-1	0	1	2	3	4	5	6	7	8	9	10	11
2	1	-24											
3	1	-12	54										
5	1	-6	9	10	-30								
7	1	-4	2	8	-5	-4	-10						
13	1	-2	-1	2	1	2	-2	0	-2	-2	1	0	0

Remarks. (1) If only the modular equation (2.2) is wanted, one can proceed more easily as follows. $j(\tau)$ is a rational function of Z of the form

$$j(\tau) = Z^{-1}(\tau) + \sum_{k=0}^{p} a_k Z^k(\tau), \tag{2.17}$$

the integers a_k being determined by the expansion of the two members of the equation at zero. Now using the invariance of j under the substitution $T\tau = -1/\tau$, we obtain

$$Z^{-1} + \sum_{0}^{p} a_k Z^k = W + \sum_{0}^{p} a_k W^{-k}, \tag{2.18}$$

where we have employed the relation

$$Z(-1/\tau) = W^{-1}(\tau),$$

obtained from (2.6). Equation (2.18) is of the same form as (2.9), from which the modular equation follows.

(2) The foregoing discussion has been carried out for a particular modular function of $\Gamma_0(p)$. However, any univalent modular function, Ψ, of $\Gamma_0(p)$, possesses a modular equation. This equation may be obtained easily from (2.2) by using the relation between Ψ and Z, which, in accordance with the univalence of both functions must be of the form

$$Z = (\alpha\Psi + \beta)/(\gamma\Psi + \delta), \qquad \alpha\delta - \beta\gamma \neq 0.$$

W is, of course, the same function of $p^{r/2}\Psi(\tau/p)$.

3. Congruences for powers of 2 and 3.

In order to prove Theorem 1 for $\alpha > 1$, we shall have to iterate equation (1.10), as was done in **5-6** of [2]. We treat the case $p = 2$ first.

In order to have (1.10) in a form suitable for iteration, we rewrite it as

$$(3.1) \qquad U_2 j = B_0 + 2^{11} \sum_{h=1}^{4} B_h 2^{8(h-1)} \Phi^h = B_0 + 2^{11} R,$$

where R is a polynomial of the form[2]

$$(3.2) \qquad R = d_1\Phi + d_2 2^8 \Phi^2 + \cdots + d_t 2^{8(t-1)} \Phi^t, \qquad (t \geq 1)$$

with integral d. Applying the operator U_2 to both sides we obtain

$$(3.3) \qquad U_2^2 j = B_0 + 2^{11} \sum_{1}^{4} B_h 2^{8(h-1)} U_2 \Phi^h = B_0 + 2^{11} U_2 R.$$

In the above equations, the B's are integers. U_2^2 is the iterate of U_2: $U_2^2 j(\tau) = U_2\{U_2 j(\tau)\}$.

We wish to have the right member of (3.3) in the same form as that of (3.1) but with a higher power of 2 multiplying the sum. We shall, in fact, prove

$$(3.4) \qquad 2^{8(h-1)} U_2 \Phi^h = 2^3 R, \qquad h = 1, 2, \cdots.$$

Theorem 1 will follow without difficulty from (3.4).

The proof of (3.4) is by induction. First, note that

$$(3.5) \qquad U_p \Phi^h = p^{-1} \sum_{\lambda=0}^{p-1} \Phi^h((\tau + \lambda)/p) = p^{-1-rh/2} \sum_{0}^{p-1} W_\lambda^h,$$

---

[2] R is not necessarily the same polynomial at each appearance.

where $W_\lambda = p^{r/2}\Phi((\tau+\lambda)/p)$ are the conjugates of the algebraic (modular) equation (2.2). Thus with

$$S_h = \sum_{0}^{p-1} W_\lambda{}^h, \tag{3.6}$$

(3.4) is equivalent to

$$S_h = 2^{4h+12}R, \qquad h = 1, 2, \cdots. \tag{3.7}$$

since $r = 24$ when $p = 2$.

The proof makes use of Newton's formulae for the sum of powers of the conjugates of an algebraic equation:[3]

$$S_h = \sum_{j=1}^{h} (-1)^{j+1} p_j S_{h-j}, \qquad h = 1, 2, \cdots, \tag{3.8}$$

with the conventions

$$p_j = 0 \text{ for } j > p,$$
$$S_0 = h,$$

where the algebraic equation is written as in (2.2). We now calculate S_1 and S_2.

From (3.8) we have

$$S_1 = p_1 \tag{3.9}$$

$$S_2 = p_1 S_1 - 2p_2 = p_1{}^2 - 2p_2. \tag{3.10}$$

From (2.3) we get

$$p_1 = 2^{16}(3\Phi + 2^8\Phi^2) = 2^{16}R \tag{3.11}$$

$$p_2 = -2^{24}\Phi = 2^{24}R, \tag{3.12}$$

where we have used the values of b_j given in the table toward the end of **2**.

The polynomials R_1

$$R_1 = 2^8 R = \sum_{j=1}^{t} d_j 2^{8j} \Phi^j \tag{3.13}$$

form a ring. In terms of R_1 we can rewrite (3.11), (3.12) as

$$p_1 = 2^8 R_1, \qquad p_2 = 2^{16} R_1. \tag{3.14}$$

Hence, from (3.9),

---

[3] The above device was first used in this connection by G. N. Watson [4].

$$S_1 = 2^{16}R, \tag{3.15}$$

$$S_2 = (2^8R_1)^2 + 2 \cdot 2^{16}R_1 = 2^{16}R_1^2 + 2^{17}R_1$$

$$= 2^{16}R_1 + 2^{17}R_1 = 2^{16}R_1 = 2^{24}R = 2^{20}R,$$

satisfying (3.7) for $h = 1, 2$.

Now let $k > 2$ and assume that (3.7) is fulfilled for $h < k$:

$$S_h = 2^{4h+12}R = 2^{4(h+1)}R_1, \qquad h = 1, 2, \cdots, k-1. \tag{3.16}$$

(3.8) gives

$$S_k = p_1S_{k-1} - p_2S_{k-2} \qquad (k > 2)$$

$$= 2^8R_1 \cdot 2^{4k}R_1 + 2^{16}R_1 : 2^{4(k-1)}R_1 \tag{3.17}$$

$$= 2^{4k+8}R_1 = 2^{4k+16}R = 2^{4k+12}R,$$

which, together with (3.15), completes the proof of (3.7), i.e., of (3.4).

A simple consequence of (3.4) is

$$U_2R = 2^3R. \tag{3.18}$$

We can now apply (3.18) to (3.3) and obtain

$$U_2^{\,2}j = B_0 + 2^{14}R. \tag{3.19}$$

By successive iteration we have

$$U_{2^m}j(\tau) = B_0 + 2^{3m+8}R, \qquad m = 1, 2, 3, \cdots. \tag{3.20}$$

Since, however, we also have

$$U_{2^m}j(\tau) = \sum_{\mu=0}^{\infty} c_{q\mu}x^{\mu}, \qquad q = 2^m \tag{3.21}$$

by iteration of (1.3), the truth of Theorem 1 follows at once for all positive integral α and $p = 2$.

The development for the case $p - 3$ is parallel. We write (1.10) as

$$U_3j = B_0 + 3^5T, \tag{3.22}$$

where T is of the form

$$T = \sum_{l=1}^{t} d_l 3^{4(l-1)}\Phi^l, \tag{3.23}$$

and the d_l are integers. We prove

$$3^{4(h-1)}U_3\Phi^h = 3^2T, \qquad h = 1, 2, \cdots, \tag{3.24}$$

or, what is equivalent

$$S_h = 3^{2h+7}T, \qquad h = 1, 2, \cdots. \tag{3.25}$$

From 2 we find without difficulty

$$\begin{aligned} p_1 &= 3^9T = 3^5T_1 \\ p_2 &= 3^{14}T = 3^{10}T_1 \\ p_3 &= 3^{18}T = 3^{14}T_1, \end{aligned} \tag{3.26}$$

where the polynomials T_1

$$T_1 = 3^4T = \sum_{l=1}^{t} d_l 3^{4l}\Phi^l \tag{3.27}$$

form a ring. From (3.8), we have

$$S_1 = p_1, \qquad S_2 = p_1{}^2 - 2p_2, \qquad S_3 = p_1{}^3 - 3p_1p_2 + 3p_3. \tag{3.28}$$

Substituting (3.26) in (3.28), we obtain

$$\begin{aligned} S_1 &= 3^9T, \\ S_2 &= (3^5T_1)^2 + 2\cdot 3^{10}T_1 = 3^{10}T_1 = 3^{14}T = 3^{11}T, \\ S_3 &= (3^5T_1)^3 + 3(3^5T_1)(3^{10}T_1) + 3\cdot 3^{14}T_1 = 3^{15}T_1 = 3^{19}T = 3^{13}T. \end{aligned} \tag{3.29}$$

Thus, (3.25) is proved for $h = 1, 2, 3$.

Now let $k > 3$, and let (3.25) be true for $h \leq k - 1$,

$$S_h = 3^{2h+7}T = 3^{2h+3}T_1, \qquad h = 1, 2, \cdots, k - 1. \tag{3.30}$$

Then, from (3.8), (3.26), and (3.30), we get

$$\begin{aligned} S_k &= p_1S_{k-1} - p_2S_{k-2} + p_3S_{k-3} \qquad (k > 3) \\ &= 3^5T_1\cdot 3^{2k+1}T_1 + 3^{10}T_1\cdot 3^{2k-1}T_1 + 3^{14}T_1\cdot 3^{2k-3}T_1 \\ &= 3^{2k+6}T_1 = 3^{2k+10}T = 3^{2k+7}T, \end{aligned} \tag{3.31}$$

as required for the proof of (3.25). Theorem 1 for $p = 3$ now follows in the same way as in the case $p = 2$.

4. Comparison with numerical results. H. S. Zuckerman has computed the coefficients c_n for $n \leq 24$ ([5]). The table which follows shows the factorization of the c_n with respect to primes < 13.

HIGHEST POWER OF p DIVIDING c_n

n \ p	2	3	5	7	11
1	2	3	0	0	0
2	11	0	1	0	0
3	1	5	1	0	0
4	14	3	0	0	0
5	3	0	2	0	0
6	13	6	0	0	1
7	0	3	1	1	0
8	17	1	3	0	0
9	2	7	1	0	0
10	12	5	3	0	0
11	1	1	0	0	1
12	16	5	1	0	0
13	3	3	1	0	0
14	14	0	0	1	0
15	0	7	2	1	0
16	20	4	1	0	0
17	3	2	1	0	0
18	11	8	1	0	0
19	1	3	3	0	0
20	15	0	2	0	0
21	7	5	0	2	0
22	13	5	1	0	1
23	2	1	1	0	0
24	19	6	0	1	0

This table serves to verify Theorem 1 and Theorems 1 and 2 of [2] for $n \leq 24$. Note that these theorems predict the exact power of the prime dividing c_n for $n = p$ and $n = p^k$ $(p < 13)$.

5. Modular functions with congruence properties.

The preceding discussion does not by any means exhaust the possibilities as regards congruence relations for Fourier coefficients of modular functions. Extensions are possible both to other moduli and modular functions other than $j(\tau)$. We close with an example of a theorem of the latter type.

THEOREM 3. *Let $F(\tau)$ be an entire modular function of $\Gamma_0(p)$, with $p \leq 11$, having the expansions*[4]

$$F(\tau) = a_{-s}x^{-s} + \cdots + a_{-1}x^{-1} + \sum_{0}^{\infty} a_n x^n, \tag{5.1}$$

---

[4] $F(\tau)$ necessarily has expansions in x at both parabolic vertices of Γ_0.

$$F(-1/p\tau) = b_{-t}x^{-t} + \cdots + b_{-1}x^{-1} + \cdots, \tag{5.2}$$

where $a_{-1}, \cdots, a_{-s}$, $b_{-1}, \cdots, b_{-t}$ *are integers, and*

$$s < p.$$

Then the Fourier coefficients a_n, *with* $n \geqq 1$, *have the congruence properties of Theorem* 1 *of this paper and Theorems* 1 *and* 2 *of* [2], *viz., if* $n \equiv 0$ $(\mathrm{mod}\ 2^a 3^b 5^c 7^d 11^e)$, *then* $a_n \equiv 0$ $(\mathrm{mod}\ 2^{3a+8} 3^{2b+3} 5^{c+1} 7^d 11^e)$ *for* $n, a, b, c, d = 1, 2, 3, \cdots$; *and* $e = 1, 2$.

Obviously $j(\tau)$ satisfies the hypotheses for all p.

Proof. By (5. 1) and (1. 3) we see that $U_pF(\tau)$ is regular at infinity, since $s < p$:

$$U_pF(\tau) = a_0 + a_px + \cdots. \tag{5.4}$$

At zero, we use (1. 4), (5. 1), (5. 4), and (5. 2), the last two with τ replaced by $p\tau$, and find

$$\begin{aligned} pU_pF(-1/p\tau) = O(1) &+ b_{-t}x^{-pt} + \cdots + b_{-1}x^{-1} + \cdots \\ &- (a_{-s}x^{-s} + \cdots + a_{-1}x^{-1}) + \cdots. \end{aligned} \tag{5.5}$$

At all interior points of the upper half-plane $U_pF(\tau)$ is regular. Moreover, it remains invariant under the substitutions of Γ_0, as may be checked by reference to Lemma 1 of [1] (p. 501). Therefore, $U_pF(\tau)$ is an entire modular function on the subgroup $\Gamma_0(p)$.

We now separate the cases $p \leqq 7$, $p = 11$, treating the former first. By fitting a polynomial in $\Phi_{p,r}$ to U_pF at $\tau' = 0$, we find in the usual way an identity of the form

$$U_pF(\tau) = \text{const.} + p^{r/2-1} \sum_{j=1}^{J} d_j p^{r(j-1)/2} \Phi^j,$$

with integral d_j, i. e.,

$$U_pF(\tau) = \text{const.} + p^{r/2-1} V, \tag{5.6}$$

where V is a polynomial of type R $(p = 2)$, T $(p = 3)$—as in **3** of this paper—or Q $(p = 5, 7)$—cf. **5** of [2]. No difficulty arises at $\tau = i\infty$ since both members of (5. 6) are regular there. The iteration of (5. 6) under the operator U_p now proceeds in exactly the same way as in **3** of this paper and **5**, **6** of [2] and establishes the congruences of the theorem when $p \leqq 7$.

For $p = 11$ we apply the methods of [2], **3**, **4** and obtain identities with integer coefficients for U_pF, $U_p{}^2F$ in terms of the functions $A_{11}(\tau)$, $C_{11}(\tau)$, which themselves have integral coefficient expansions. The congruences for $p = 11, 11^2$ are then immediate consequences.

Added in proof. One additional value of c_n appears in the literature, namely c_{25}, on p. 491 of D. H. Lehmer's paper, "Properties of the coefficients of the modular invariant $J(\tau)$," *American Journal of Mathematics*, vol. 64 (1942), pp. 488-502. c_{25} has the factorization $2^2 \cdot 3^3 \cdot 5^3 \cdot \lambda$, where $p < 13$ does not divide λ. This value further illustrates the remark at the close of **4.**

Theorem 3 can be extended to the case $e = 3$. See addendum at the end of [2].

NEW YORK CITY.

BIBLIOGRAPHY.

[1] J. Lehner, "Ramanujan identities involving the partition function for the moduli 11^a," *American Journal of Mathematics*, vol. 65 (1943), pp. 492-520.

[2] ———, "Divisibility properties of the Fourier coefficients of the modular invariant $j(\tau)$," *American Journal of Mathematics*, vol. 71 (1949), pp. 136-148.

[3] H. Rademacher, "The Ramanujan identities under modular substitutions," *Transactions of the American Mathematical Society*, vol. 51 (1942), pp. 609-636.

[4] G. N. Watson, "Ramanujans Vermutung über Zerfällungsanzahlen," *Journal für die reine und angewandte Mathematik*, vol. 179 (1938), pp. 97-128.

[5] H. S. Zuckerman, "The computation of the smaller coefficients of $J(\tau)$," *Bulletin of the American Mathematical Society*, vol. 45 (1939), pp. 917-919.

DIOPHANTINE EQUATIONS IN ALGEBRAIC NUMBER FIELDS.*

By L. G. PECK.

Consider a system of algebraic equations,

$$f_1(x_1,\cdots,x_q)=0,\qquad \cdots\cdots \qquad f_h(x_1,\cdots,x_q)=0, \tag{1}$$

where f_i is a homogeneous polynomial of degree m_i with coefficients belonging to the ring J of integers of a given finite algebraic extension K of the rational field. The main object of this paper is to prove the following result.

THEOREM 1. *If K is totally complex, then for every system of degrees $m_1,\cdots,m_h$ and every positive rational integer m there exists a positive number $Q=Q(m;m_1,\cdots,m_h;K)$ such that whenever $q\geqq Q$ the system of equations* (1) *has an m-dimensional linear manifold of solutions in J.*

Using Theorem C of [I],[1] one sees that Theorem 1 is equivalent to the following

THEOREM 2. *If K is totally complex, then for every positive rational integer m there exists a positive number $Q=Q(m;K)$ such that whenever $q\geqq Q$ the equation*

$$\alpha_1\lambda_1{}^m+\cdots+\alpha_q\lambda_q{}^m=0 \qquad (\alpha_1,\cdots,\alpha_q\,\varepsilon\, J) \tag{2}$$

has non-trivial solutions in J (the trivial solution being $\lambda_1=\cdots=\lambda_q=0$).

It should be remarked at once that the requirement that K be totally complex is also necessary for the conclusions in Theorems 1 and 2, since in the case that K has real conjugates there exist forms in arbitrarily many variables one of whose real conjugates is positive definite. Nevertheless, it will turn out to be possible to obtain a generalization of Theorem 2 for non-totally complex fields by adding the requirement that the form $\alpha_1x_1{}^m+\cdots+\alpha_qx_q{}^m$ be indefinite (cf. Theorem 3 below).

The method which I use in the proof was developed by C. L. Siegel [II, III] for the treatment of Waring's problem in algebraic number fields.

---

* Received November 17, 1948.

[1] Numbers in square brackets indicate papers in the list of references at the end.

Siegel's calculations for the form $x_1^m + \cdots + x_q^m$ apply with only slight modifications to the form $\alpha_1 x_1^m + \cdots + \alpha_q x_q^m$, the only essential difference arising in the proof that the "singular series" has a positive lower bound. I wish to express here my deep gratitude to Professor Siegel for his invaluable advice in the preparation of this paper.

1. Notation and plan of proof. Let K be an algebraic field of degree n over the rationals. $K^{(1)}, \cdots, K^{(r)}$ are the real conjugates and $K^{(r+1)}, \cdots, K^{(r+2s)}$ the complex conjugates of K. Here $r + 2s = n$. The complex conjugates are so arranged that, for $r + 1 \leqq k \leqq r + s$, the fields $K^{(k)}$ and $K^{(k+s)}$ are obtained from one another by interchanging $i (= \sqrt{-1})$ and $-i$.

Elements of K are denoted by small Greek letters; their conjugates by the same letters with superscripts. The norm and trace are defined by $N(\alpha) = \alpha^{(1)} \cdots \alpha^{(n)}$, $S(\alpha) = \alpha^{(1)} + \cdots + \alpha^{(n)}$, the same definition holding also for Greek letters which are not necessarily elements of K.

If $\omega_1, \cdots, \omega_n$ is a basis of J, then the transposed inverse $(\rho_l^{(k)})$ of the matrix $(\omega_l^{(k)})$ yields a basis $\rho_1, \cdots, \rho_n$ of the ideal $\mathfrak{d}^{-1}$, where $\mathfrak{d}$ is the ramification ideal. For the absolute value D of the discriminant of K it follows that $D^{\frac{1}{2}} = |\det(\omega_l^{(k)})|$, $D = N(\mathfrak{d})$. It will be important to recall the following property of $\mathfrak{d}$:

$\delta \in \mathfrak{d}^{-1}$ *if and only if* $S(\omega_l \delta) =$ *a rational integer for* $l = 1, 2, \cdots, n$.

Introducing real variables x_k, y_k, y_{kl} $(k = 1, \cdots, n;\ l = 1, 2, \cdots)$, define

$$\begin{aligned} \xi &= \rho_1 x_1 + \cdots + \rho_n x_n, \qquad \eta = \omega_1 y_1 + \cdots + \omega_n y_n, \\ \eta_l &= \omega_1 y_{1l} + \cdots + \omega_n y_{nl} \qquad (l = 1, 2, \cdots). \end{aligned} \tag{3}$$

The points $(x_1, \cdots, x_n)$, $(y_1, \cdots, y_n)$, *etc.*, are to be regarded as variable points in an n-dimensional Euclidean space X. The conjugates of ξ, η, η_l are obtained by inserting the superscript $^{(k)}$ above each of the Greek letters in (3). All relations involving Greek letters without superscripts (except those with N or S) will henceforth stand for the n relations obtained by inserting superscripts as above.

A homogeneous polynomial $f(z_1, \cdots, z_q)$ with coefficients in K will be called (in)definite if each of the real polynomials $f^{(k)}(z_1, \cdots, z_q)$ $(k = 1, \cdots, r)$ is (in)definite. Thus, in totally complex fields, the requirement of (in)definiteness is no restriction at all. It must be remarked, however, that in fields with real conjugates there are polynomials which are neither definite nor indefinite (*e. g.*, if $\theta^2 = 2$, one conjugate of $z^2 - \theta w^2$ is indefinite while the other is definite). The generalization of Theorem 2 mentioned in the introduction may now be stated.

THEOREM 3. *If* $q = 1 + \max [4m^{2n+3}, (2^{m-1} + n)mn]$ *and if the polynomial* $\alpha_1 z_1{}^m + \cdots + \alpha_q z_q{}^m$ *is indefinite, then* (2) *has non-trivial solutions in* J.

Note that q depends only on the degree n of K. Hence, in Theorems 1 and 2, $Q(m; m_1, \cdots; K)$ and $Q(m; K)$ may be replaced by $Q(m; m_1, \cdots; n)$ and $Q(m; n)$.

If $\alpha_j = 0$, then (2) has the solution $\lambda_j = 1$, $\lambda_k = 0$ $(k \neq j)$; hence it will be assumed throughout that $\alpha_k \neq 0$ $(k = 1, \cdots, q)$.

The first step in the proof is to obtain an expression for the number $\mathfrak{N}(T)$ of solutions of (2) satisfying $|\lambda_l| < T$ $(l = 1, \cdots, q)$. To this end, using the notation $1^u = e^{2\pi i u}$ introduce the function

$$f(x; \alpha) = \sum_{|\lambda| < T} 1^{S(\alpha\lambda^m\xi)} \qquad (\alpha \in J; \alpha \neq 0),$$

and define

$$f(x) = f(x; \alpha_1) \cdots f(x; \alpha_q).$$

From the fact that the function $1^{S(\mu\xi)}$ has, when μ is an integer, period 1 in each of the variables $x_1, \cdots, x_n$, it follows that

$$\int_E 1^{S(\mu\xi)} dx = \begin{cases} 0 & (\mu \neq 0), \\ 1 & (\mu = 0), \end{cases} \qquad (dx = dx_1 \cdots dx_n)$$

where the domain of integration is the unit cube

$$E\colon\ 0 < x_k < 1 \qquad (k = 1, \cdots, n).$$

Hence

$$\mathfrak{N}(T) = \int_E f(x)\, dx. \tag{4}$$

In what follows, a slight modification of Siegel's generalized Farey dissection will be applied to the cube E and the contributions of the major and minor arcs will then be estimated. In this way it will be shown that

$$\mathfrak{N}(T) = T^{n(q-m)}\sigma I + o(T^{n(q-m)}) \tag{5}$$

where the "singular series" σ and the "singular integral" I will turn out to be positive and independent of T. In (5) and throughout the paper, the symbols o and O refer to the passage to the limit $T \to \infty$.

2. The Farey-Siegel dissection. Since Theorem 3 is trivial for $m = 1$, assume $m \geqq 2$ and let

$$a = (2^{m-1} + n)^{-1}, \qquad t = T^{1-a}.$$

Let

$$c = \max_{\substack{k=1,\ldots,r+s\\ l=1,\ldots,q}} |\alpha_l^{(k)}|,$$

and suppose T is sufficiently large to ensure $T^{2a} > 2cD^{1/n}$.

For each γ in K the ideal $\mathfrak{a}_\gamma$ is defined by the relation $\mathfrak{a}_\gamma^{-1} = (1, \gamma\mathfrak{d})$, and for any positive real c the domain $B(\gamma, c)$ is the set of points $x \varepsilon X$ which satisfy

$$N[\max(T^m |\xi - \gamma|, c^{-1})] \leqq t^n N(\mathfrak{a}_\gamma^{-1}).$$

(The domain B_γ in [III] is here denoted by $B(\gamma, 1)$.) Clearly, $B(\gamma, c)$ is empty if $N(\mathfrak{a}_\gamma) > (ct)^n$; moreover, $B(\gamma, c)$ *and* $B(\gamma', c)$ *have no common point when* $\gamma \neq \gamma'$, for if x were a common point of the two regions and

$$\max(T^m |\xi - \gamma|, c^{-1}) = t\sigma^{-1}, \qquad \max(T^m |\xi - \gamma'|, c^{-1}) = t\sigma'^{-1},$$

then

$$\sigma \leqq ct, \qquad \sigma' \leqq ct, \qquad N(\mathfrak{a}_\gamma \mathfrak{a}_{\gamma'}) \leqq N(\sigma\sigma'),$$

and

$$\begin{aligned} |\gamma - \gamma'| &\leqq |\xi - \gamma| + |\xi - \gamma'| \leqq T^{-m} t(\sigma^{-1} + \sigma'^{-1}) \\ &\leqq 2cT^{2-2a-m}\sigma^{-1}\sigma'^{-1}, \\ N((\gamma - \gamma')\mathfrak{a}_\gamma \mathfrak{a}_{\gamma'}) &\leqq N(|\gamma - \gamma'|) N(\sigma\sigma') \\ &\leqq (2cT^{2-2a-m})^n < D^{-1}; \end{aligned}$$

this would mean that the integral ideal $(\gamma - \gamma')\mathfrak{a}_\gamma \mathfrak{a}_{\gamma'}\mathfrak{d}$ has norm less than 1, which is impossible.

Because of the periodicity of $f(x)$ it is clear that whenever $\gamma \equiv \gamma'$ $(\mathrm{mod}\ \mathfrak{d}^{-1})$,

$$\int_{B(\gamma,c)} f(x)\,dx = \int_{B(\gamma',c)} f(x)\,dx.$$

Consequently, if Γ is a complete set of modulo $\mathfrak{d}^{-1}$ incongruent numbers γ satisfying $N(\mathfrak{a}_\gamma) \leqq (ct)^n$, and if $X_0(c)$ is the subset of X consisting of those points which (for the particular c under consideration) do not lie in any $B(\gamma, c)$, then (4) may be written in the form

$$\mathfrak{R}(T) = \sum_\gamma \int_{B(\gamma,c)} f(x)\,dx + \int_{X_0(c) \cap E} f(x)\,dx, \tag{6}$$

where γ (here and in the sequel) runs over the finite set Γ.

3. The minor arcs. The last integral in (6) will be estimated in the three lemmata of this section.

LEMMA 1. *If* α *is an integer,* $\alpha \neq 0$, *then*

$$N(\mathfrak{a}_{\alpha\gamma}^{-1}) \leqq N(|\alpha|) N(\mathfrak{a}_{\gamma}^{-1}).$$

Proof. $\mathfrak{a}_{\alpha\gamma}^{-1}$ consists of all numbers of the form $\lambda + \alpha\gamma\delta$ where $\lambda \varepsilon J$, $\delta \varepsilon \mathfrak{d}$; on the other hand, $\alpha\mathfrak{a}_{\gamma}^{-1}$ consists of all numbers of the form $\alpha(\lambda + \gamma\delta)$ where again $\lambda \varepsilon J$, $\delta \varepsilon \mathfrak{d}$. Hence $\mathfrak{a}_{\alpha\gamma}^{-1} \mid \alpha\mathfrak{a}_{\gamma}^{-1}$ and the lemma follows.

LEMMA 2. *For the integer* α $(\neq 0)$ *let*

$$\xi' = \alpha\xi = \rho_1 x'_1 + \cdots + \rho_n x'_n.$$

If $|\alpha| \leqq c$ *and* $x \varepsilon X_0(c)$, *then* $x' \varepsilon X_0(1)$.

Proof. Let $\gamma' = \alpha\gamma$. Since

$$N[\max(T^m |\xi - \gamma|, c^{-1})] > t^n N(\mathfrak{a}_{\gamma}^{-1})$$

for every γ in K, it follows that

$$\begin{aligned} N[\max(T^m |\xi' - \gamma'|, 1)] &\geqq N[\max(T^m |\alpha| \, |\xi - \gamma|, |\alpha| c^{-1})] \\ &> t^n N(|\alpha|) N(\mathfrak{a}_{\gamma}^{-1}) \geqq t^n N(\mathfrak{a}_{\gamma'}^{-1}) \end{aligned}$$

for every γ' in K. This proves Lemma 2.

LEMMA 3. *If* $x \varepsilon X_0(c)$ *and* $|\alpha| \leqq c$ *then*

$$f(x;\alpha) = o(T^{n(1-m/q)}).$$

Proof. By Lemma 2, $x' \varepsilon X_0(1)$. Now the proof of Lemma 8 in [III] may be used with the following modifications: x is to be replaced by x', ξ by ξ', and the set $\mathfrak{T}$ by the set of integers λ satisfying $|\lambda| < T$.

Lemma 3 makes it possible to write (6) in the form

$$\mathfrak{N}(T) = \sum_{\gamma} \int_{B(\gamma,c)} f(x)\,dx + o(T^{n(q-m)}). \tag{7}$$

4. The major arcs.

LEMMA 4. *For any integer* α $(\neq 0)$ *set*

$$G_a(\gamma) = N(\mathfrak{a}_{\gamma}^{-1}) \sum_{\mu \bmod \mathfrak{a}_{\gamma}} 1^{S(a\mu^m\gamma)}$$

where μ *runs over a complete set of incongruent integers modulo the ideal* $\mathfrak{a}_\gamma$. *If* $x \in B(\gamma, c)$, *then*

$$f(x;\alpha) = G_\alpha(\gamma) \int_{|\eta|<T} 1^{S[\alpha\eta^m(\xi-\gamma)]} dy + O(T^{n-a}),$$

where η *is defined by* (3) *and* $dy = dy_1 \cdots dy_n$.

Proof. Since $N[\max(T^m|\xi-\gamma|, c^{-1})] \leqq t^n N(\mathfrak{a}_\gamma^{-1})$ and $N(\mathfrak{a}_\gamma) \leqq (ct)^n$, one can determine positive numbers $\theta^{(1)}, \cdots, \theta^{(n)}$ with $\theta^{(k)} = \theta^{(k+s)}$ $(k = r+1, \cdots, r+s)$ such that

$$\theta \max(T^m|\xi-\gamma|, c^{-1}) \leqq tD^{1/2n}, \qquad N(\theta) = \mathrm{D}^{1/2} N(\mathfrak{a}_\gamma).$$

By Minkowski's theorem, the ideal $\mathfrak{a} = \mathfrak{a}_\gamma$ contains a number $\bar{\alpha}$ satisfying $0 < |\bar{\alpha}| \leqq \theta$. Then $\bar{\alpha}\mathfrak{a}^{-1} = \mathfrak{b}$ is integral and $N(\mathfrak{b}) \leqq D^{1/2}$, so that $\mathfrak{b}$ belongs to a finite set depending only on K. The basis $\beta_1, \cdots, \beta_n$ of $\mathfrak{b}^{-1}$ gives rise to a basis $\bar{\alpha}_k = \bar{\alpha}\beta_k$ $(k = 1, \cdots, n)$ of $\mathfrak{a}$; and, since $\beta_k = O(1)$, it follows that $\bar{\alpha}_k = O(\theta)$.

Clearly

$$f(x;\alpha) = \sum_{\mu \bmod \mathfrak{a}} 1^{S(\alpha\mu^m\gamma)} \sum_{\substack{|\lambda+\mu|<T \\ \mathfrak{a}|\lambda}} 1^{S[\alpha(\lambda+\mu)^m(\xi-\gamma)]}, \tag{8}$$

where μ runs over a complete set of integral residues modulo $\mathfrak{a}$ and λ runs over all elements of $\mathfrak{a}$ which satisfy $|\lambda+\mu| < T$. Introduce the variable point z in X and let $\zeta = \bar{\alpha}_1 z_1 + \cdots + \bar{\alpha}_n z_n$. If $\lambda = g_1\bar{\alpha}_1 + \cdots + g_n\bar{\alpha}_n$ with rational integers $g_1, \cdots, g_n$ then the cube $E(\lambda)$ may be defined as the set of points z satisfying $g_k \leqq z_k < g_k + 1$ $(k = 1, \cdots, n)$. It follows that when $z \in E(\lambda)$,

$$\zeta - \lambda = O(\theta) = O(t),$$

and

$$\begin{aligned} &\alpha(\xi-\gamma)[(\zeta+\mu)^m - (\lambda+\mu)^m] \\ &\qquad = \alpha(\zeta-\lambda)(\xi-\gamma)O(|\zeta+\mu|^{m-1} + |\lambda+\mu|^{m-1}) \\ &\qquad = \alpha\theta(\xi-\gamma)O(T^{m-1}) = \alpha t T^{-m} O(T^{m-1}) = O(T^{-a}); \end{aligned}$$

hence

$$1^{S[\alpha(\lambda+\mu)^m(\xi-\gamma)]} = \int_{E(\lambda)} 1^{S[\alpha(\zeta+\mu)^m(\xi-\gamma)]} dz + O(T^{-a}). \tag{9}$$

For a fixed μ the conjugates of any λ which occurs in the inner sum in (8) all lie in fixed intervals of length $O(T)$. Hence the number of these λ is $N(\mathfrak{a}^{-1})O(T^n)$, and therefore the error term which results when (9) is summed with respect to λ is $N(\mathfrak{a}^{-1})O(T^{n-a})$.

For real u, let $F(u)$ be the domain defined by the inequality $|\zeta+\mu| < T+u\theta$; then the volume of $F(u)$ is

$$(10) \qquad V(u) = \begin{cases} 2^{r+s}\pi^s D^{-1/2} N(\mathfrak{a}^{-1}) N(T+u\theta) & (T+u\theta>0), \\ 0 & \text{(otherwise)}. \end{cases}$$

There exists a positive constant $A = O(1)$ such that

$$F(-A) \subset \sum_{\lambda} E(\lambda) \subset F(A);$$

and, on the other hand,

$$F(-A) \subset F(0) \subset F(A).$$

Since (10) implies that

$$V(A) - V(-A) = N(\mathfrak{a}^{-1}) T^{n-1} O(\theta) = N(\mathfrak{a}^{-1}) O(T^{n-a}),$$

it follows from (9) that

$$(11) \qquad \sum_{\substack{|\lambda+\mu|<T \\ \mathfrak{a}|\lambda}} 1^{S[a(\lambda+\mu)^m(\xi-\gamma)]} = \int_{F(0)} 1^{S[a(\zeta+\mu)^m(\xi-\gamma)]} dz + N(\mathfrak{a}^{-1}) O(T^{n-a}).$$

Using (3), substitute $\zeta+\mu=\eta$ in the above integral. Since

$$\partial(z_1,\cdots,z_n)/\partial(y_1,\cdots,y_n) = N(\mathfrak{a}^{-1}),$$

the lemma is a consequence of (8) and (11).

LEMMA 5. *For $N(\mathfrak{a}_\gamma)\to\infty$, the estimate*

$$G_a(\gamma) = N(\mathfrak{a}_\gamma^{-1/m}) O(1)$$

holds.

This will be proved in **7**.

LEMMA 6.

$$\int_{|\eta|<T} 1^{S[a\eta^m(\xi-\gamma)]} dy = N[\min(1, T^{-1}|\alpha(\xi-\gamma)|^{-1/m})] \cdot O(T^n).$$

This follows in the same way as equations (80)-(83) in [III]; the details of the proof are therefore omitted.

To obtain an estimate of the first term on the right side of (7), set $G(\gamma) = G_{a_1}(\gamma)\cdots G_{a_q}(\gamma)$ and $du_l = dy_{1l}\cdots dy_{nl}$ $(l=1,\cdots,q)$. Then lemmata 4, 5, 6 imply

$$(12)\quad f(x) = G(\gamma)\int_{|\eta_1|<T}\cdots\int_{|\eta_q|<T} 1^{S[(\alpha_1\eta_1^m+\cdots+\alpha_q\eta_q^m)(\xi-\gamma)]}du_1\cdots du_q$$
$$+\,O(T^{q(n-a)}) + N(\mathfrak{a}_\gamma^{(1-q)/m})N[\min(1, T^{1-q}\,|\,c(\xi-\gamma)\,|^{\,(1-q)/m})]O(T^{qn-a}).$$

Moreover,

$$(13)\qquad \int_{B(\gamma,c)} N[\min(1, T^{1-q}\,|\,c(\xi-\gamma)\,|^{\,(1-q)/m})]dx$$
$$\leqq T^{-ma}\int_X N[\min(1,\,|\,c\xi\,|^{\,(1-q)/m})]dx = O(T^{-mn}),$$

and (since $q-1>2m$)

$$(14)\qquad \sum_\gamma N(\mathfrak{a}_\gamma^{(1-q)/m}) \leqq \sum_\mathfrak{a} N(\mathfrak{a}^{1+(1-q)/m}) = O(1).$$

Using the obvious estimate $\sum_\gamma \int_{B(\gamma,c)} dx \leqq 1$ and the fact that $a(q-1)\geqq mn$, it now follows from (7), (12), (13), and (14) that

$$\mathfrak{R}(T) =$$
$$\sum_\gamma G(\gamma)\int_{B(\gamma,c)}\int_{|\eta_1|<T}\cdots\int_{|\eta_q|<T} 1^{S[\omega(\xi-\gamma)]}du_1\cdots du_q + o(T^{n(q-m)}),$$

where $\omega = \alpha_1\eta_1^m + \cdots + \alpha_q\eta_q^m$. As in [III, p. 336], it is now possible to replace the integration domain $B(\gamma, c)$ by the entire space X with an error of $o(T^{n(q-m)})$. Thus,

$$(15)\qquad \mathfrak{R}(T) = \sigma(T)I(T) + o(T^{n(q-m)})$$

where

$$(16)\qquad \sigma(T) = \sum_\gamma G(\gamma)$$

and

$$(17)\qquad I(T) = \int_X\int_{|\eta_1|<T}\cdots\int_{|\eta_q|<T} 1^{S(\omega\xi)}du_1\cdots du_q dx.$$

6. The singular integral. If the substitutions $\eta_l \to T\eta_l$ $(l=1,\cdots,q)$ and $\xi \to T^{-m}\xi$ are made in the integral (17), there results

$$I(T) = T^{n(q-m)}I(1).$$

It will be shown in this section that the singular integral $I(1) = I$, which is clearly independent of T, is positive. The following two cases require separate consideration: (i) $r>0$ and $m\equiv 0 \pmod 2$; (ii) all others.

Let u_l denote the point $(y_{1l},\cdots,y_{nl})$ in X, and consider the set $\mathfrak{E}$ of points $(u_1,\cdots,u_q)$ in an nq-dimensional Euclidean space defined by the inequalities

$$|\eta_l| < 1 \quad (l = 1,\cdots,q), \qquad 0 < \eta_q^{(k)} < 1 \quad (k = 1,\cdots,r),$$
$$-\pi/m < \arg \eta_q^{(k)} < \pi/m \quad (k = r+1,\cdots,r+s) \text{ in case (i)};$$
$$|\eta_l| < 1 \quad (l = 1,\cdots,q), \qquad -\pi/m < \arg \eta_q^{(k)} < \pi/m$$
$$(k = r+1,\cdots,r+s) \text{ in case (ii)}.$$

Further, let

$$\Phi(\xi) = w \int_{\mathfrak{E}} 1^{S(\omega\xi)}\, du_1 \cdots du_q$$

where $w = 2^r m^s$ in case (i), $w = m^s$ in case (ii), and ω has the same meaning as in **5**. If $J(\mu) = \int_X \Phi(\xi) 1^{-S(\mu\xi)} dx$ (the integral being absolutely convergent by the reasoning at the end of **5**), then clearly $I = J(0)$.

For a fixed $\omega = \omega_1 t_1 + \cdots + \omega_n t_n$, let $\mathfrak{E}(\omega)$ be the set of points $(u_1,\cdots,u_{q-1})$ in an $n(q-1)$-dimensional Euclidean space for which u_q can be found satisfying $\omega = \alpha_1\eta_1{}^m + \cdots + \alpha_q\eta_q{}^m$ and $(u_1,\cdots,u_q) \,\varepsilon\, \mathfrak{E}$. Then every $(u_1,\cdots,u_q) \,\varepsilon\, \mathfrak{E}$ gives rise to a unique ω and every ω and $(u_1,\cdots,u_{q-1}) \,\varepsilon\, \mathfrak{E}(\omega)$ gives rise to a unique $u_q \,\varepsilon\, X$. Therefore, since

$$\partial(y_{1q},\cdots,y_{nq})/\partial(t_1,\cdots,t_n) = N(m^{-1}\,|\,\alpha_q\,|^{-1}\,|\,\eta_q\,|^{1-m}),$$

it follows that

$$\Phi(\xi) = w \int_X \Psi(\omega) 1^{S(\omega\xi)}\, dt,$$

where

$$\Psi(\omega) = \int_{\mathfrak{E}(\omega)} N(m^{-1}\,|\,\alpha_q\,|^{-1}\,|\,\eta_q\,|^{1-m})\, du_1 \cdots du_{q-1}. \tag{18}$$

It is clear that $\Psi(\omega)$ vanishes outside of a certain bounded region, *e.g.*, $\Psi(\omega) = 0$ when $|\,\omega^{(k)}\,| > |\,\alpha_1{}^{(k)}\,| + \cdots + |\,\alpha_q{}^{(k)}\,|$ for some k. The continuity of $\Psi(\omega)$ is, moreover, a consequence of the following calculation.

In case (ii), set

$$\alpha_l{}^{(k)}(\eta_l{}^{(k)})^m = v_{kl} \qquad (k = 1,\cdots,r;\ l = 1,\cdots,q-1);$$
$$|\,\alpha_l{}^{(k)}\,|^2\,|\,\eta_l{}^{(k)}\,|^{2m} = v_{kl}, \qquad \arg \alpha_l{}^{(k)}(\eta_l{}^{(k)})^m = \phi_{kl}$$
$$(k = r+1,\cdots,r+s;\ l = 1,\cdots,q-1).$$

Then, for $l = 1,\cdots,q-1$, the Jacobian of the y's which have second subscript l with respect to the v's and ϕ's which have second subscript l is $D^{-\frac{1}{2}}N(m^{-1}\,|\,\alpha_l\,|^{-1}\,|\,\eta_l\,|^{1-m})$, and

$$\Psi(\omega) = D^{(1-q)/2}m^{-nq}\,|\,N(\alpha_1)\cdots N(\alpha_q)\,|^{-1/m}\prod_{k=1}^{r}F_k(\omega^{(k)})\prod_{k=r+1}^{r+s}H_k(\omega^{(k)}),$$

where, for $k = 1,\cdots,r$,

$$F_k(\omega^{(k)}) = \int \,|\,v_1\cdots v_{q-1}(\omega^{(k)} - v_1 - \cdots - v_{q-1})\,|^{(1-m)/m}dv_1\cdots dv_{q-1},$$

the integral being extended over the domain

$$\begin{aligned} &|\,v_l\,| < |\,\alpha_l^{(k)}\,| && (l = 1,\cdots,q-1)\\ &|\,\omega^{(k)} - v_1 - \cdots - v_{q-1}\,| < |\,\alpha_q^{(k)}\,|\,; \end{aligned}$$

and for $k = r+1,\cdots,r+s$,

$$H_k(\omega^{(k)}) = \int \,|\,v_1\cdots v_{q-1}(\omega^{(k)} - v_1^{\frac{1}{2}}e^{i\phi_1} - \cdots - v_{q-1}^{\frac{1}{2}}e^{i\phi_{q-1}})^2\,|^{(1-m)/m} \cdot dv_1\cdots dv_{q-1}d\phi_1\cdots d\phi_{q-1},$$

the integral being extended over the domain

$$\begin{aligned} &0 < v_l < |\,\alpha_l^{(k)}|^2, \quad -\pi < \phi_l < \pi \qquad (l = 1,\cdots,q-1),\\ &|\,\omega^{(k)} - v_1^{\frac{1}{2}}e^{i\phi_1} - \cdots - v_{q-1}^{\frac{1}{2}}e^{i\phi_{q-1}}\,| < |\,\alpha_q^{(k)}|\,. \end{aligned}$$

In case (i), the domain $\mathfrak{E}(\omega)$ may be split into $2^{r(q-1)}$ domains, in each of which the quantities $\eta_l^{(k)}$ $(k = 1,\cdots,r;\ l = 1,\cdots,q-1)$ are all of constant sign. In any one of these domains, new variables may be introduced by the formulae

$$\begin{aligned} &|\,\alpha_l^{(k)}\,|(\eta_l^{(k)})^m = v_{kl} \qquad (k = 1,\cdots,r;\ l = 1,\cdots,q-1)\,;\\ &|\,\alpha_l^{(k)}|^2\,|\,\eta_l^{(k)}\,|^{2m} = v_{kl}, \qquad \arg\alpha_l^{(k)}(\eta_l^{(k)})^m = \phi_{kl}\\ &\qquad\qquad (k = r+1,\cdots,r+s;\ l = 1,\cdots,q-1), \end{aligned}$$

and, as before, one obtains

$$\Psi(\omega) = 2^{r(q-1)}D^{(1-q)/2}m^{-nq}\,|\,N(\alpha_1)\cdots N(\alpha_q)\,|^{-1/m}\prod_{k=1}^{r}F_k(\omega^{(k)})\prod_{k=r+1}^{r+s}H_k(\omega^{(k)},$$

where, for $k = 1,\cdots,r$,

$$F_k(\omega^{(k)}) = \int \,|\,v_1\cdots v_{q-1}(\omega^{(k)} - (\operatorname{sgn}\alpha_1^{(k)})v_1 - \cdots - (\operatorname{sgn}\alpha_{q-1}^{(k)})v_{q-1})\,|^{(1-m)/m}\cdot dv_1\cdots dv_{q-1},$$

the integral extended over the domain

$$(19)\qquad \begin{aligned} &0 < v_l < |\alpha_l^{(k)}| && (l = 1, \cdots, q-1),\\ &0 < [\omega^{(k)} - (\operatorname{sgn}\alpha_1^{(k)})v_1 - \cdots - (\operatorname{sgn}\alpha_{q-1}^{(k)})v_{q-1}]/\alpha_q^{(k)} < 1; && \end{aligned}$$

and $H_k(\omega^{(k)})$ has the same meaning as in case (ii).

Classical considerations show that $F_k(\omega^{(k)})$ and $H_k(\omega^{(k)})$ are in both cases continuous. Hence $\Psi(\omega)$ is continuous and Fourier's theorem shows that $J(\omega) = w\Psi(\omega)$. Thus, $I = w\Psi(0)$.

It is clear that the quantities $H_k(0)$ $(k = r+1, \cdots, r+s)$ are positive; also in case (ii) $F_k(0) > 0$ $(k = 1, \cdots, r)$. In case (i), the requirement that $\alpha_1 z_1^m + \cdots + \alpha_q z_q^m$ be indefinite means that for each fixed k in the interval $1 \leqq k \leqq r$ not all of the quantities $\alpha_1^{(k)}, \cdots, \alpha_q^{(k)}$ have the same sign. Hence the domain (19) with $\omega^{(k)} = 0$ has a positive volume and again $F_k(0) > 0$ $(k = 1, \cdots, r)$. Thus, in any case, $I = w\Psi(0) > 0$.

7. The singular series. The following four lemmata have as their immediate objective the proof of lemma 5, but will be used later in this section for the estimate of the quantity $\sigma(T)$ in (16).

LEMMA 7. *If* $(\mathfrak{a}_\gamma, \mathfrak{a}_{\gamma'}) = 1$ *then* $\mathfrak{a}_{\gamma+\gamma'} = \mathfrak{a}_\gamma \mathfrak{a}_{\gamma'}$ *and*

$$G_a(\gamma + \gamma') = G_a(\gamma) G_a(\gamma').$$

Proof. The hypotheses $(\mathfrak{a}_\gamma, \gamma\mathfrak{d}\mathfrak{a}_\gamma) = 1$, $(\mathfrak{a}_\gamma, \mathfrak{a}_{\gamma'}) = 1$ imply

$$1 = (\mathfrak{a}_\gamma, \gamma\mathfrak{d}\mathfrak{a}_\gamma\mathfrak{a}_{\gamma'}) = (\mathfrak{a}_\gamma, (\gamma + \gamma')\mathfrak{d}\mathfrak{a}_\gamma\mathfrak{a}_{\gamma'}).$$

Similarly,

$$1 = (\mathfrak{a}_{\gamma'}, (\gamma + \gamma')\mathfrak{d}\mathfrak{a}_\gamma\mathfrak{a}_{\gamma'})$$

and therefore

$$1 = (\mathfrak{a}_\gamma\mathfrak{a}_{\gamma'}, (\gamma + \gamma')\mathfrak{d}\mathfrak{a}_\gamma\mathfrak{a}_{\gamma'}),$$

i. e.,

$$\mathfrak{a}_\gamma^{-1}\mathfrak{a}_{\gamma'}^{-1} = (1, (\gamma + \gamma')\mathfrak{d}) = \mathfrak{a}_{\gamma+\gamma'}^{-1}.$$

If λ, λ' are integers such that

$$\mathfrak{a}_\gamma \mid \lambda, \qquad \mathfrak{a}_{\gamma'} \mid \lambda', \qquad (\lambda, \mathfrak{a}_{\gamma'}) = 1, \qquad (\lambda', \mathfrak{a}_\gamma) = 1,$$

then

$$\begin{aligned} N(\mathfrak{a}_\gamma\mathfrak{a}_{\gamma'}) G_a(\gamma + \gamma') &= \sum_{\mu \bmod \mathfrak{a}_\gamma\mathfrak{a}_{\gamma'}} 1^{S[a\mu^m(\gamma+\gamma')]}\\ &= \sum_{\mu \bmod \mathfrak{a}_\gamma} \sum_{\mu' \bmod \mathfrak{a}_{\gamma'}} 1^{S[a(\lambda'\mu+\lambda\mu')^m(\gamma+\gamma')]}\\ &= \sum_{\mu \bmod \mathfrak{a}_\gamma} \sum_{\mu' \bmod \mathfrak{a}_{\gamma'}} 1^{S[a(\mu^m\gamma+\mu'^m\gamma')]} = N(\mathfrak{a}_\gamma) N(\mathfrak{a}_{\gamma'}) G_a(\gamma) G_a(\gamma'), \end{aligned}$$

which proves the lemma.

Let $\mathfrak{p}$ be a prime ideal, and for any integer α let $P(\alpha)$ be the rational integer for which $\mathfrak{p}^{P(\alpha)} \mid \alpha$ and $\mathfrak{p}^{P(\alpha)+1} \nmid \alpha$. In what follows, π is any integer satisfying $P(\pi) = 1$.

Lemma 8. *If* $\mathfrak{a}_\gamma = \mathfrak{p}^l$, $l \geqq 2P(m) + 2$, *then*

$$G_1(\gamma) = N(\mathfrak{p}^{-1}) G_1(\pi^m \gamma).$$

Proof.

$$\begin{aligned} N(\mathfrak{p}^l) G_1(\gamma) &= \sum_{\mu \bmod \mathfrak{p}^l} 1^{S(\mu^m \gamma)} \\ &= \sum_{\mu \bmod \mathfrak{p}^{l-P(m)-1}} \quad \sum_{\nu \bmod \mathfrak{p}^{P(m)+1}} 1^{S[(\mu+\pi^{l-P(m)-1}\nu)^m \gamma]} \\ &= \sum_{\mu \bmod \mathfrak{p}^{l-P(m)-1}} \quad \sum_{\nu \bmod \mathfrak{p}^{P(m)+1}} 1^{S[(\mu^m + m\mu^{m-1}\pi^{l-P(m)-1}\nu)\gamma]}. \end{aligned}$$

The inner sum equals $N(\mathfrak{p}^{P(m)+1}) 1^{S(\mu^m\gamma)}$ if $\mathfrak{p}^l \mid m\mu^{m-1}\pi^{l-P(m)-1}$, *i. e.*, if $\mathfrak{p} \mid \mu$; otherwise this sum vanishes. Hence

$$\begin{aligned} N(\mathfrak{p}^{l-P(m)-1}) G_1(\gamma) &= \sum_{\substack{\mu \bmod \mathfrak{p}^{l-P(m)+1} \\ \mathfrak{p} \mid \mu}} 1^{S(\mu^m\gamma)} \\ &= \sum_{\mu \bmod \mathfrak{p}^{l-P(m)-2}} 1^{S(\mu^m \pi^m \gamma)} = N(\mathfrak{p}^{l-P(m)-2}) G_1(\pi^m\gamma). \end{aligned}$$

The last step is clear when $m \geqq P(m) + 2$. If $m < P(m) + 2$, it must be observed that

$$\begin{aligned} N(\mathfrak{p}^{l-m}) G_1(\pi^m\gamma) &= \sum_{\mu \bmod \mathfrak{p}^{l-m}} 1^{S(\mu^m\pi^m\gamma)} \\ &= \sum_{\mu \bmod \mathfrak{p}^{l-P(m)-2}} \quad \sum_{\nu \bmod \mathfrak{p}^{P(m)+2-m}} 1^{S[(\mu+\pi^{l-P(m)-2}\nu)^m \pi^m\gamma]} \\ &= N(\mathfrak{p}^{P(m)+2-m}) \sum_{\mu \bmod \mathfrak{p}^{l-P(m)-2}} 1^{S(\mu^m\pi^m\gamma)}, \end{aligned}$$

so that the step in question is again valid. This proves the lemma.

Lemma 9. *If* $\mathfrak{a}_\gamma = \mathfrak{p}^l$, $l \geqq 2P(m\alpha) + 2$, *and the rational integer* k *satisfies* $l - (k-1)m \geqq 2P(m\alpha) + 2$, *then*

$$G_\alpha(\gamma) = N(\mathfrak{p}^{-k}) G_\alpha(\pi^{km}\gamma).$$

Proof. Since $\mathfrak{a}_{\alpha\gamma} = \mathfrak{p}^{l-P(\alpha)}$, it is clear that $G_\alpha(\gamma) = G_1(\alpha\gamma)$, and the desired result follows by repeated application of Lemma 8.

Lemma 10. *If* $\mathfrak{a}_\gamma = \mathfrak{p}$ *and* $N(\mathfrak{p})$ *is sufficiently large (specifically, if* $N(\mathfrak{p}^{(m-2)/2m}) \geqq m - 1$ *and* $\mathfrak{p} \nmid \alpha$*), then*

$$| G_\alpha(\gamma) | \leqq N(\mathfrak{p}^{-1/m}).$$

Proof. Let Ξ be the set of characters of the multiplicative group Λ of integers modulo $\mathfrak{p}$. If $\delta = \alpha\gamma$, then $\mathfrak{a}_\delta = \mathfrak{p}$ and, for any non-principal character $\chi \varepsilon \Xi$,

$$\Big|\sum_{\lambda \varepsilon \Lambda} 1^{S(\lambda\delta)}\bar{\chi}(\lambda)\Big|^2 = \sum_{\lambda \varepsilon \Lambda}\sum_{\mu \varepsilon \Lambda} 1^{S[(\lambda-\mu)\delta]}\bar{\chi}(\lambda)\chi(\mu)$$
$$= \sum_{\lambda \varepsilon \Lambda}\sum_{\nu \varepsilon \Lambda} 1^{S[(1-\nu)\lambda\delta]}\chi(\nu) = \sum_{\nu \varepsilon \Lambda}\sum_{\lambda \bmod \mathfrak{p}} 1^{S[(1-\nu)\lambda\delta]}\chi(\nu) = N(\mathfrak{p}).$$

Hence

$$|N(\mathfrak{p})G_a(\gamma)| = \Big|1 + \sum_{\lambda \varepsilon \Lambda} 1^{S(\lambda^m\delta)}\Big|$$
$$= \Big|1 + (N(\mathfrak{p})-1)^{-1} \sum_{\lambda \varepsilon \Lambda}\sum_{\mu \varepsilon \Lambda}\sum_{\chi \varepsilon \Xi} 1^{S(\mu\delta)}\bar{\chi}(\mu)\chi(\lambda^m)\Big|$$
$$= \Big|1 + \sum_{\substack{\chi \varepsilon \Xi \\ \chi^m = 1}}\sum_{\mu \varepsilon \Lambda} 1^{S(\mu\delta)}\bar{\chi}(\mu)\Big| \leqq [(m, N(\mathfrak{p})-1)-1]N(\mathfrak{p}^{1/2})$$
$$\leqq (m-1)N(\mathfrak{p}^{1/2}) \leqq N(\mathfrak{p}^{(m-1)/m}),$$

which proves the lemma.

Using Lemma 9 with the largest permissible value of k together with Lemma 10, one sees easily that when $\mathfrak{a}_\gamma = \mathfrak{p}^l$ and $N(\mathfrak{p})$ is sufficiently large, $|G_a(\gamma)| \leqq N(\mathfrak{p}^{-l/m})$. On the other hand, if $\mathfrak{p}$ is one of the finite number of prime ideals whose norms are not "sufficiently large" then Lemma 9 shows that $G_a(\gamma) = N(\mathfrak{p}^{-l/m})O(1)$. These facts, together with the multiplicative property of $G_a(\gamma)$ proved in Lemma 7 lead at once to Lemma 5.

An immediate consequence of Lemma 5 is that the singular series

$$\sigma = \sum_{\delta \bmod \mathfrak{d}^{-1}} G(\delta)$$

is majorized by $\sum_{\delta \bmod \mathfrak{d}^{-1}} N(\mathfrak{a}_\delta^{-q/m}) < \sum_{\mathfrak{a}} N(\mathfrak{a}^{1-(q/m)})$ (the latter sum running over all integral ideals $\mathfrak{a}$) and is therefore convergent (since $q > 2m$). Hence $\sigma(T) = \sigma + o(1)$, and Theorem 3 will follow when it is shown that $\sigma > 0$.

Let $H(\mathfrak{a}) = \sum_{\substack{\delta \bmod \mathfrak{d}^{-1} \\ (1,\delta\mathfrak{d}) = \mathfrak{a}^{-1}}} G(\delta)$. Lemma 7 shows that $H(\mathfrak{a}\mathfrak{b}) = H(\mathfrak{a})H(\mathfrak{b})$ whenever $(\mathfrak{a}, \mathfrak{b}) = 1$, and thus, because of the absolute convergence of all series involved, the Euler factorization of $\sigma = \sum_{\mathfrak{a}} H(\mathfrak{a})$ is given by

$$\sigma = \prod_{\mathfrak{p}} \sigma_{\mathfrak{p}} \qquad (\mathfrak{p} \text{ runs over all prime ideals}), \tag{20}$$

where

$$\sigma_{\mathfrak{p}} = \sum_{l=0}^{\infty} H(\mathfrak{p}^l). \tag{21}$$

Since the product (20) is convergent, it is now sufficient to show that, for every prime ideal $\mathfrak{p}$, $\sigma_{\mathfrak{p}} > 0$.

Lemma 9 shows that $H(\mathfrak{p}^l) = N(\mathfrak{p}^{m-q})H(\mathfrak{p}^{l-m})$ whenever $l \geqq 2P(m\alpha_j) + 2$ $(j = 1, \cdots, q)$ and $l \geqq m$. Hence writing

$$\Omega = \max\,[2P(m\alpha_1) + 1, \cdots, 2P(m\alpha_q) + 1, m - 1],$$

it is clear that

$$(22) \quad \sigma_{\mathfrak{p}} = \sum_{l=0}^{\Omega-m} H(\mathfrak{p}^l) + (1 + N(\mathfrak{p}^{m-q}) + N(\mathfrak{p}^{2(m-q)}) + \cdots) \sum_{l=\Omega-m+1}^{\Omega} H(\mathfrak{p}^l),$$

where the first term is to be deleted when $\Omega = m - 1$.

The partial sum

$$\sum_{l=0}^{k} H(\mathfrak{p}^l) = N(\mathfrak{p}^{-qk}) \sum_{\substack{\gamma \bmod \mathfrak{d}^{-1} \\ 1 \mid \gamma\mathfrak{d}\mathfrak{p}^k}} \sum_{\mu_1 \bmod \mathfrak{p}^k} \cdots \sum_{\mu_q \bmod \mathfrak{p}^k} 1^{S[(\alpha_1\mu_1{}^m + \dots + \alpha_q\mu_q{}^m)\gamma]}$$

is easily computed with the aid of the formula

$$\sum_{\substack{\delta \bmod \mathfrak{d}^{-1} \\ 1 \mid \delta\mathfrak{d}\mathfrak{p}^k}} 1^{S(\lambda\delta)} = \begin{cases} N(\mathfrak{p}^k) \text{ if } \mathfrak{p}^k \mid \lambda, \\ \quad 0 \quad \text{otherwise,} \end{cases} \qquad (1 \mid \lambda).$$

The result is

$$\sum_{l=0}^{k} H(\mathfrak{p}^l) = N(\mathfrak{p}^{(1-q)k})M(k)$$

where $M(k)$ is the number of solutions of the congruence

$$(23) \qquad \alpha_1\mu_1{}^m + \cdots + \alpha_q\mu_q{}^m \equiv 0 \pmod{\mathfrak{p}^k}.$$

Consequently, (22) yields

$$(24) \qquad \sigma_{\mathfrak{p}} = (1 - N(\mathfrak{p}^{m-q}))^{-1}[N(\mathfrak{p}^{(1-q)\Omega})M(\Omega) \\ - N(\mathfrak{p}^{(1-q)\Omega+qm-q})M(\Omega - m)].$$

Let $M_0(k)$ be the number of primitive solutions of (23), *i. e.*, solutions satisfying $(\mu_1, \cdots, \mu_q, \mathfrak{p}) = 1$. Any non-primitive solution of (23) may be written in the form

$$(25) \qquad \mu_1 \equiv \lambda_1\pi, \cdots, \mu_q \equiv \lambda_q\pi \pmod{\mathfrak{p}^k},$$

and (25) is a solution of (23) if and only if (for $m \leqq k$)

$$\alpha_1\lambda_1{}^m + \cdots + \alpha_q\lambda_q{}^m \equiv 0 \pmod{\mathfrak{p}^{k-m}}.$$

Since $M(k - m)$ counts the number of sets of λ's modulo $\mathfrak{p}^{k-m}$, whereas in (25) they must be counted modulo $\mathfrak{p}^{k-1}$, it follows that

$$M(k) - M_0(k) - N(\mathfrak{p}^{q(m-1)})M(k - m) \qquad (k \geqq m);$$

and therefore, from (24),

$$\sigma_{\mathfrak{p}} = (1 - N(\mathfrak{p}^{m-q}))^{-1} N(\mathfrak{p}^{(1-q)\Omega}) M_0(\Omega).$$

To prove $\sigma_{\mathfrak{p}} > 0$, it therefore remains only to show that $M_0(\Omega) > 0$. The concluding argument will actually yield $M_0(k) > 0$ for every k. If $P(\alpha_j) \geqq k$, then (23) has the solution $\mu_j = 1$, $\mu_l = 0$ $(l \neq j)$. Hence it will be assumed in what follows that $k > P(\alpha_j)$ $(j = 1, \cdots, q)$.

An equivalence relation for integers in K is defined by the statement, $\alpha \sim \beta$ whenever $P(\alpha) \equiv P(\beta) \pmod{m}$. Since there are m equivalence classes, it is clear that if q' is the smallest integer not less than q/m then at least q' of the integers $\alpha_1, \cdots, \alpha_q$ belong to the same class. Suppose without loss of generality that $\alpha_1 \sim \alpha_2 \sim \cdots \sim \alpha_{q'}$; $P(\alpha_j) = h_0 + mh_j$ $(j = 1, \cdots, q')$, $0 \leqq h_0 < m$. Then integers β_j can be found satisfying $\alpha_j \equiv \pi^{h_0 + mh_j} \beta_j \pmod{\mathfrak{p}^k}$, $(\beta_j, \mathfrak{p}) = 1$ $(j = 1, \cdots, q')$.

Consider now a second equivalence relation for integers of K defined by the statement, $\alpha \approx \beta$ whenever $(\alpha, \mathfrak{p}) = (\beta, \mathfrak{p}) = 1$ and $\alpha \equiv \lambda^m \beta \pmod{\mathfrak{p}^k}$ for some integer λ. If N_k is the number of solutions of the congruence $\lambda^m \equiv 1 \pmod{\mathfrak{p}^k}$, then $(N(\mathfrak{p}^k) - N(\mathfrak{p}^{k-1}))/N_k$ is the number of incongruent m-th power residues modulo $\mathfrak{p}^k$ which are relatively prime to $\mathfrak{p}$, and therefore N_k is the number of equivalence classes under $\approx$.

To estimate N_k, observe first that $N_1 \leqq m$, since the multiplicative group modulo $\mathfrak{p}$ is cyclic. It is also clear that $N_k \leqq N(\mathfrak{p}) N_{k-1}$ $(k = 2, 3, \cdots)$, whence

$$N_k \leqq N(\mathfrak{p}^{k-1}) m \qquad (k = 1, 2, \cdots). \tag{26}$$

If $k \geqq 2P(m) + 1$ and $\lambda^m \equiv 1 \pmod{\mathfrak{p}^k}$, then $\mu \equiv \lambda \pmod{\mathfrak{p}^{k-P(m)}}$ implies $\mu \equiv \lambda + \pi^{k-P(m)} \nu \pmod{\mathfrak{p}^{k+1}}$, whence $\mu^m \equiv (\lambda + \pi^{k-P(m)} \nu)^m \equiv \lambda^m \equiv 1 \pmod{\mathfrak{p}^k}$ independently of ν, whereas, writing $\lambda^m - 1 \equiv \pi^k \alpha \pmod{\mathfrak{p}^{k+1}}$ and $m \equiv \pi^{P(m)} \beta$ $\pmod{\mathfrak{p}^{k+1}}$ with $(\beta, \mathfrak{p}) = 1$, the congruence $1 \equiv \mu^m \equiv \lambda^m + m\pi^{k-P(m)} \lambda^{m-1} \nu$ $\pmod{\mathfrak{p}^{k+1}}$ is seen to be satisfied only when $\alpha + \beta\lambda^{m-1}\nu \equiv 0 \pmod{\mathfrak{p}}$. Thus the number of solutions of $\mu^m \equiv 1 \pmod{\mathfrak{p}^k}$, $\mu \equiv \lambda \pmod{\mathfrak{p}^{k-P(m)}}$ is $N(\mathfrak{p}^{P(m)})$ and the number of solutions of $\mu^m \equiv 1 \pmod{\mathfrak{p}^{k+1}}$, $\mu \equiv \lambda \pmod{\mathfrak{p}^{k-P(m)}}$ is also $N(\mathfrak{p}^{(m)})$; in other words, $N_{k+1} = N_k$ when $k \geqq 2P(m) + 1$. Taken together with (26), this information yields the inequality

$$N_k \leqq N(\mathfrak{p}^{2P(m)}) m \leqq N(m^2) m = m^{2n+1}.$$

If q'' is the smallest integer not less than q'/m^{2n+1}, the last inequality shows that at least q'' of the integers $\beta_1, \cdots, \beta_{q'}$ belong to a single equivalence class under $\approx$. Suppose without loss of generality that $\beta_1 \approx \cdots \approx \beta_{q''}$ and $h_1 \leqq \cdots \leqq h_{q''}$. Determine integers γ_j all relatively prime to $\mathfrak{p}$ such that $\gamma_j{}^m\beta_j \equiv \beta_{q''} \pmod{\mathfrak{p}^k}$ $(j = 1, \cdots, q'' - 1)$, and set

$$\mu_j = \begin{cases} \pi^{h_{q''}-h_j}\gamma_j\nu_j & (j = 1, \cdots, q'' - 1), \\ 1 & (j = q''), \\ 0 & (j = q'' + 1, \cdots, q). \end{cases} \tag{27}$$

Then (23) becomes

$$\alpha_{q''}(\nu_1{}^m + \cdots + \nu_{q''-1}{}^m + 1) \equiv 0 \pmod{\mathfrak{p}^k},$$

which certainly holds if $\nu_1, \cdots, \nu_{q''-1}$ are rational integers satisfying

$$\nu_1{}^m + \cdots + \nu_{q''-1}{}^m \equiv -1 \pmod{p^{k'}} \tag{28}$$

where p is the rational prime divisible by $\mathfrak{p}$ and $k' = k - P(\alpha_{q''})$.

Since $q \geqq 4m^{2n+3} + 1$, it follows that $q' \geqq 4m^{2n+2} + 1$ and $q'' \geqq 4m + 1$. But it is known that every rational integer is congruent modulo an arbitrary rational prime power to a sum of $4m$ m-th powers. Hence (28) has indeed a solution. This may be substituted in (27) to obtain a solution of (23) which, since $\mu_{q''} = 1$, is certainly primitive.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[I] R. Brauer, "A note on systems of homogeneous algebraic equations," *Bulletin of the American Mathematical Society*, vol. 51 (1945), p. 749.

[II] C. L. Siegel, "Generalization of Waring's problem to algebraic number fields," *American Journal of Mathematics*, vol. 66 (1944), p. 122.

[III] ———, "Sums of m-th powers of algebraic integers," *Annals of Mathematics*, vol. 46 (1945), p. 313.

SOME DIOPHANTINE ASPECTS OF MODULAR FUNCTIONS.*

I. ESSENTIAL SINGULARITIES.

By Harvey Cohn.

1. Introduction. The modular group is a natural invariant involved in many problems in diophantine approximation. Some very basic properties of real numbers x are preserved under the transformations of the modular group: $x' = (ax + b)/(cx + d)$, where a, b, c, d are integers satisfying the relation $ad - bc = 1$. To cite a few of these properties, we might mention the rationality of x and the boundedness of the partial quotients in the simple continued fraction for x.

Although any real function of the real variable x, invariant under the transformations of the modular group, is either constant or totally discontinuous, there are functions of a complex variable x, analytic in the upper half plane, and possessing the desired invariance. Aside from the constant function, these so-called modular functions all have essntial singularities on the real axis. Our problem is then that of investigating the influence of the diophantine character of a real number x on the *essential* singularity of certain modular functions at x.

Hardy and Littlewood [3], [4] [1] carried out such an investigation for the purpose of setting bounds on certain exponential sums (given by the Jacobi theta functions on their boundary circle) by studying these functions near the singularity in question. Conversely, Wintner [7] analyzed the behavior of the Dedekind η-function by considering the coefficients in its power series expansion. Although we present a discussion of the singularities of certain modular functions as related to the singularities of the terms of the infinite series representing them, our objectives will be primarily in the realm of function theory, since we shall, for instance, demonstrate function-theoretical properties in the large on the basis of the behavior of modular functions near singularities.

---

* Received April 21, 1948; revised November 20, 1948; presented to the American Mathematical Society, June 19, 1948. Based on the author's doctoral dissertation (Harvard, 1948).

[1] Numbers in brackets refer to bibliography at the end of the article.

2. Outline. We choose as the particular modular functions, the Poincaré theta functions, $f(z)$, defined by the property

$$f((az+b)/(cz+d)) = (cz+d)^{2m}f(z), \tag{1}$$

a property which generalizes the behavior of Jacobi theta functions. We shall use as our starting point some well known results of Poincaré [6], which we summarize in the next section.

Poincaré theta functions interest us because they have the advantage of being easily obtained as infinite series whose individual terms often suggest the behavior at essential singularities. For instance, we can take the Eisenstein series,

$$F_{2m}(z) = \sum_{(c,d)\neq(0,0)} (cz+d)^{-2m} \tag{2}$$

where (c, d) runs over all integer pairs except $(0, 0)$, and where m is an integer $\geqq 2$. This is a Poincaré theta function in which each individual term becomes singular only at one rational point, $z = p/q$. At such a point, $F_{2m}(z)$ behaves like $(z - p/q)^{-2m}$ if z approaches p/q in a Stolz neighborhood[2] at p/q. Rational points, furthermore, are characterized by this behavior, as we shall see in subsequent paragraphs.

If we were interested in the behavior of modular functions at $z = \theta$, where θ is a real number whose continued fraction expansion possesses bounded partial quotients, we would find it natural to consider a sum of the type $\Sigma(Az^2 + Bz + C)^{-2m}$, where the summation goes over all forms *equivalent*, in the classical sense, to a given indefinite form. This, too, would generate a Poincaré theta function. The individual terms of its series have singularities only at certain real quadratic irrationals, which assuredly have continued fractions with bounded (in fact with ultimately periodic) partial quotients. We shall single out for special attention the more conglomerate Poincaré theta function,

$$\sum_{ad-bc=1} (cz^2 + (a+d)z + b)^{-2m} = G_{2m}(z), \tag{3}$$

where the summation extends over all "substitutions" in the modular group. This function behaves at real irrational numbers θ whose continued fractions have partial quotients like $O(z - \theta)^{-2m}$ as z approaches θ in a Stolz neighborhood. In fact when $m = 6$, and only then, the behavior is $\asymp (z-\theta)^{-2m}$.[3]

---

[2] A Stolz neighborhood of θ, in the present context, is a region of the z-plane with θ as a boundary point and contained in a sector of the form $y/|x-\theta| > \lambda$, where $\lambda > 0$.

[3] We write $f(z) \sim g(z)$ or $f(z) \asymp g(z)$, when $f(z)/g(z)$ approaches a constant or remains bounded away from 0 and infinity, respectively.

As we shall note later on, this behavior is also, in a sense, characteristic of this type of irrational.

It should be pointed out, however, that the term by term analysis by which we guess at the behavior near essential singularities can not always be readily supported. True, we can justify the term by term analysis with regard to the rational singularities of $F_{2m}(z)$, but we shall not even attempt by a term by term analysis to explain how $G_{2m}(z)$ can vanish identically, as it does when $m = 2, 3, 4, 5,$ or 7. We shall, therefore, have to rely on function-theoretical as well as analytic methods, expressing Poincaré theta functions in terms of the better known elliptic modular function $J(z)$.

We shall see that the results applying to $F_{2m}(z)$ and $G_{2m}(z)$ are just special cases of results for general theta functions, which we shall establish presently. We shall also see that the asymptotic behavior, to a certain extent, can describe the function and the singularity.

3. Preliminary results on theta functions. We shall now collect some results on the convergence of Poincaré theta functions. These results are either proved in the literature or are immediate consequences of such results.

The modular group defined in **1** is generated by the substitutions $z \rightarrow z + 1$ and $z \rightarrow -1/z$. It has a *fundamental domain* $\mathfrak{D}$ consisting of the region $\mathcal{R}^*$ of the upper half plane defined by the inequalities $|z| > 1$ and $|\Re z| < \frac{1}{2}$, together with the boundary points of $\mathcal{R}^*$ identified according to the two transformations $z \rightarrow z + 1$ and $z \rightarrow -1/z$. Every point z of the upper half plane is equivalent, modulo the modular group, to a single point in the fundamental domain $\mathfrak{D}$. (The point at infinity, being a boundary point, is excluded from the upper half plane.)

We define the elliptic modular function, with Dedekind, as the function $J(z)$ which maps the fundamental domain onto the whole z-plane so that $J(\infty) = \infty$, $J(i) = 1$, and $J(\frac{1}{2}(1 + i\sqrt{3})) = 0$. The mapping function, $J(z)$ by virtue of the shape of the fundamental domain, will take on the value unity twice at $z = i$, and will have a triple zero at $z = \frac{1}{2}(1 + i\sqrt{3})$. As $\Im z$ becomes infinite, $J(z)$ likewise will behave like

$$(\exp - 2\pi i z)(k_0 + k_1 \exp 2\pi i z + k_2 \exp 4\pi i z + \cdots).$$

The exact value of k_0, namely $1/1728$, can not be immediately inferred from the shape of the fundamental domain, but we shall concern ourselves only with the more evident fact that $k_0 \neq 0$. From its definition, $J(z)$ is invariant under the modular group. It can be further shown that any single-valued function of z, invariant under the modular group, is a single-valued function of $J(z)$.

A Poincaré theta function is defined as a single-valued function $f(z)$ regular in $\mathfrak{D}$ with the exception of a finite number of poles, and satisfying the transformation formula (1). We call m the *order* of the Poincaré theta function. Thus $J'(z)$ is of order 1, and $f(z)J'(z)^{-m}$ is a single-valued function of $J(z)$. In particular, if it is known that $f(z)$ has no singularities in the upper half plane, it easily follows that the only poles of $f(z)J'(z)^{-m}$ are those induced by the zeros of $J'(z)$, and these can be easily accounted for. Thus we can write

$$f(z) = J'(z)^m J(z)^{-m_0}(J(z)-1)^{-m_1}P(J(z)), \tag{4}$$

where $P(w)$ is an entire function of w. Suitable values of m_0 and m_1 are given in the following table [6; 599]:[4]

$m \equiv 0 \pmod 6$	$m_0 = 2m/3$	$m_1 = m/2$	$m_2 = m/6$
1	$2(m-1)/3$	$(m-1)/2$	$(m-7)/6$
2	$(2m-1)/3$	$m/2$	$(m-2)/6$
3	$2m/3$	$(m-1)/2$	$(m-3)/6$
4	$2(m-1)/3$	$m/2$	$(m-4)/6$
5	$(2m-1)/3$	$(m-1)/2$	$(m-5)/6$

The Poincaré theta functions, having the period 1, can always be expanded into a Laurent series in $\exp 2\pi i z$ valid for $\Im z$ positive and large. In some cases the series is finite in one direction so that

$$f(z) = (\exp - 2\pi i r z)(k'_0 + k'_1 \exp 2\pi i z + \cdots),$$

where $k'_0 \neq 0$ and r is an integer. (This occurs, for instance, when $f(z)$ is bounded as $\Im z \to \infty$.) Then we say $f(z)$ is of *degree r*. Thus if $f(z)$ has no poles the function $P(w)$ introduced in equation (4) is a polynomial of degree $r + m_2$, where m_2 is given in the above table. For other function-theoretical aspects of theta functions, too numerous to recall here, we refer to Poincaré's papers. We are principally interested in the tables just presented.

4. Preliminary results on theta series.

We are now ready to consider some examples of theta functions expressed as infinite series. We consider the series,

---

[4] The integer m_2 will be explained presently.

$$(2) \qquad F_{2m}(z) = \sum_{(c,d)\neq(0,0)} (cz+d)^{-2m}$$

and

$$(5) \qquad \Theta_{2m}(z) = \sum_{ad-bc=1} (cz+d)^{-2m}(\omega + (az+b)/(cz+d))^{-1},$$

where $m \geqq 2$ and the first and second sums are over all integer sets that satisfy the conditions stated under the summation. In either case it is apparent that we are dealing with a theta function if we can assume absolute convergence, since a rearrangement of terms shows that (1) is satisfied by $F_{2m}(z)$ and $\Theta_{2m}(z)$.

In a region $\mathfrak{R}$, bounded away from the real axis and having only a finite projection on the real axis, $F_{2m}(z)$ converges absolutely and uniformly, by well known estimates.

If, in the series for $\Theta_{2m}(z)$ as given in equation (5), we sum first over all a, b satisfying $a \equiv \alpha \pmod{c}$ and $b \equiv \beta \pmod{c}$, α, β, c and d being fixed, then we obtain

$$(5\text{a}) \qquad \Theta_{2m}(z) = \sum_{ad-bc=1}^{*} (cz+d)^{-2m}\,\pi\cot\pi(\omega + (az+b)/(cz+d)),$$

where the asterisk means that a and b run independently over complete sets of residues modulo c. We now assume once and for all that $\Im\omega \geq 0$; thus we avoid poles. The series (5) with appropriate grouping of terms, or series (5a), converges absolutely and uniformly in the region $\mathfrak{R}$ mentioned above.

If we differentiate the series (5) term by term we obtain a series in which the convergence is now absolute and uniform, without grouping of terms, in the region $\mathfrak{R}$. Thus we define

$$(6) \qquad \begin{aligned} G_{2m}(z,\omega) &= -(d/d\omega)^{2m-1}\Theta_{2m}(z)/(2m-1)! \\ &= \sum_{ad-bc=1} (cz\omega + az + d\omega + b)^{-2m}. \end{aligned}$$

Since $G_{2m}(z,z)$ is precisely $G_{2m}(z)$, we see how to obtain the latter theta series from $\Theta_{2m}(z)$.

5. Non-vanishing of theta functions. We are now faced with the by no means trivial problem of deciding if our theta series vanish identically. Without such a decision, the problem of asymptotic behavior may be meaningless.

The case of $F_{2m}(z)$ is easily disposed of. If z approaches infinity along the imaginary axis, by the uniformity of convergence of its series, $F_{2m}(z)$ must approach the part of its sum consisting of constant terms, which makes for a limit of $2\zeta(2m)$. Hence $F_{2m}(z)$ is not identically zero for any integer $m \geq 2$.

Next consider $\Theta_{2m}(z)$ as written in the form (5a). By the uniformity of convergence again, we can let z become infinite along a line parallel to the imaginary axis and find that $\Theta_{2m}(z)$ approaches $-2\pi i$, (since only the terms with $a = \pm 1$, $b = 0$, $c = 0$, $d = \pm 1$ then remain in (5a)). But now, in expansion (4), $m_2 = 0$ or P is a constant when $\Theta_{2m}(z)$ is expressed in terms of the elliptic modular function for $m = 2, 3, 4, 5$ or 7. The same is true for $F_{2m}(z)$. Hence comparing the values at infinity, we find $\Theta_{2m}(z) = -F_{2m}(z) i\pi\zeta(2m)^{-1}$. Thus, despite appearances, $\Theta_{2m}(z)$ does not depend on ω, and its derivative with respect to ω must be identically zero; in particular so are $G_{2m}(z, \omega)$ and $G_{2m}(z)$, for these small values of m.

To see the dependence on ω of $\Theta_{2m}(z)$ for all other values of m, we consider the expansion of this function into powers of $\exp 2\pi i z$ and $\exp 2\pi i\omega$, which leads to the following formula [6; 615]:

$$-\Theta_{2m}(z)(4\pi i)^{-1} - F_{2m}(z)(4\zeta(2m))^{-1} = \sum_{n,k=1} A_{nk} \exp 2\pi i(nz + k\omega).$$

$$A_{nk} = \delta_{nk} + (-)^m 2\pi \sum_{c=1}^{\infty} (n/k)^{m-\frac{1}{2}} c^{-1} J_{2m-1}(4\pi(nk)^{\frac{1}{2}}/c) \cdot \sum_{aa'\equiv 1\,(c)} \exp 2\pi i(ka + na')/c. \tag{7}$$

Here δ_{nk} is the Kronecker symbol, and $J_{2m-1}(z)$ is the Bessel function of indicated order. As noted by Poincaré, when $m = 2, 3, 4, 5$ or 7, each A_{nk} is zero, which leads to some rather remarkable identities. We are more directly concerned with showing that $A_{11} \neq 0$ if $m = 6$ or $m \geq 8$, whence none of the derivatives of $\Theta_{2m}(z)$ with respect to ω are identically zero. To see this, we need only estimate the Bessel functions occurring in (7) to show that, when $n = k = 1$, the second term in the expression for A_{nk} is much less than δ_{11} $(=1)$. It is sufficient, for our purposes, to know that the exponential sum in (7) is less than c. The tedious details of the estimations can be omitted, since they simply consist of taking enough terms of the power series for $J_{2m-1}(z)$ about the origin, so that the remaining terms in the expansion will constitute an alternating series. We find that two terms suffice in this case. We easily make the estimate $A_{11} = 1 + (-)^m 2\pi J_{2m-1}(4\pi)$ + error of less than .01. Furthermore, $J_{11}(4\pi) = .291\cdots$, $J_{13}(4\pi) = .159\cdots$ (rather close to $(2\pi)^{-1}$), and $J_{2m-1}(4\pi) < .05$ when $m \geq 8$. This finally establishes the dependence on ω of Θ_{2m} for $m = 6$ and $m \geq 8$. The non-vanishing of these A_{11} even shows that in $G_{2m}(z, \omega)$ the leading term $\exp 2\pi i(z + \omega)$ has a non-zero coefficient.

Collecting the results of the above calculation, we find that $F_{2m}(z)$ is of order m and degree 0 for $m \geq 2$, while $G_{2m}(z)$ is of order $2m$ and of degree -2, for $m = 6$, or $m \geq 8$. The expansions in (4) become

$$F_{2m}(z) = J'(z)^m J(z)^{-m_0}(J(z) - 1)^{-m_1}(J(z)^{m_2} + k_1''J(z)^{m_2-1} + \cdots + k_{m_2}'') \cdot 2\zeta(2m)(-2\pi i)^{-m} \tag{8a}$$

and, by symmetry,

$$G_{2m}(z, \omega) = J'(z)^m J(z)^{-m_0}(J(z) - 1)^{-m_1}P_{m_2-1}(J(z)) \cdot J'(\omega)^m J(\omega)^{-m_0}(J(\omega) - 1)^{-m_1}P_{m_2-1}(J(\omega)) \tag{8b}$$

where P_{m_2-1} is a polynomial of indicated degree. Setting $\omega = z$, we find for $m = 6$ and $m \geq 8$

$$G_{2m}(z) = [J'(z)^m J(z)^{-m_0}(J(z) - 1)^{-m_1}P_{m_2-1}(J(z))]^2. \tag{8c}$$

The expression for $G_{12}(z)$ in the equation (8c) is of some interest since it is proportional to the square of $J'(z)^6 J(z)^{-4}(J(z) - 1)^{-3}$, which in turn is proportional to the twenty-fourth power of $\eta(z)$, the Dedekind η-function. To see the last statement, we combine the result [2], $J(z) = F_4(z)^3\eta(z)^{-24}/1728$ with the expansion for $F_4(z)$ in equation (8a).

We see from the table in the last section that only Poincaré theta functions of order $0, \pm 6, \pm 12, \cdots$, can be devoid of zeros and poles. It is also clear that $G_{12}(z)^{\frac{1}{2}k}$ is such a function when k is an integer. Thus *the theta function* $f(z)$ *has no zeros or poles if and only if* $F(z) = G_{12}(z)^{\frac{1}{2}k} \exp g(J(z))$, *where* $g(w)$ *is an entire function of* w. Hence among functions of *finite degree,* the only theta functions devoid of zeros and poles are $G_{12}(z)^{\frac{1}{2}k}$ times a constant.

6. Asymptotic behavior and term by term estimation. The function $F_{2m}(z) = \Sigma(cz + d)^{-2m}$, $m \geq 2$, has the advantage of permitting a term by term analysis of its behavior at rational points. Specifically, let z approach p/q in a Stolz neighborhood, or let $z - p/q = \rho_1 + i\rho_2$ where ρ_1 and ρ_2 are real, $\rho_2 > 0$, $|\rho_2/\rho_1| \geq \lambda$, (a parameter of slope of the Stolz neighborhood), and $\rho_1{}^2 + \rho_2{}^2 \to 0$. We break up $F_{2m}(z)$ into two parts, one part $F^{(1)}$ containing just those terms which become infinite at $z = p/q$, and another part $F^{(2)}$ consisting of the remaining terms.

Then $F^{(1)}$ is the summation restricted to those terms where $c/d = -q/p$, or $F^{(1)} = (z - p/q)^{-2m}2q^{-2m}\zeta(2m)$.

It remains to estimate the terms where $c/d \neq -q/p$. Setting $z = p/q + \rho_1 + i\rho_2$, we find

$$F^{(2)} = \sum_{cp-dq\neq 0} (cp/q + d + c\rho_1 + ic\rho_2)^{-2m}.$$

It is majorized by its series of absolute values,

$$q^{2m} \sum [(cp + dq + c\rho_1 q)^2 + (cq\rho_2)^2]^{-m}.$$

Now the terms may be grouped as follows: Since $(p, q) = 1$, $cp + dq$ takes on each non-zero integer n for a family of values (c, d) of the form $c = c_n + \mu q$, $d = d_n - \mu p$, where (c_n, d_n) is a special solution to $cp + dq = n$ determined by the condition $0 \leq c_n < q$, and μ takes on all integral values. We may now take our majorant as follows:

$$(9a) \qquad |F^{(2)}| \leq q^{2m} \sum_{n=-\infty}^{\infty}{}' \sum_{\mu=-\infty}^{\infty} \{[n + \rho_1 q^2(\mu + c_n/q)]^2 + [\rho_2 q^2(\mu + c_n/q)]^2\}^{-m},$$

where $n \neq 0$.

We consider the summation on n for μ fixed. The quantity $A_n = n + \rho_1 q^2(\mu + c_n/q)$ has the difference $A_{n+k} - A_n = k + \rho_1 q(c_{n+k} - c_n)$, which differs from k by less than $\frac{1}{2}$, if $\rho_1 q^2 < \frac{1}{2}$, as we shall now suppose. Hence the sum

$$S_1 = q^{2m} \sum_{n=-\infty}^{\infty} \sum_{\mu=-\infty}^{\infty} [(|n| - \tfrac{1}{2})^2 + (\rho_2 q^2 \mu)^2]^{-m}$$

majorizes the sum in (9a) except for those terms of (9a) where, for given μ, n takes on such a value as to make $|A_n| < \frac{1}{2}$. But clearly, if so, $|\rho_1 q^2(\mu + c_n/q)| > |n| - \frac{1}{2}$ and $\rho_2 q^2(\mu + c_n/q) > |\rho_2/\rho_1|(|n| - \frac{1}{2})$. Adding this contribution to S_1 we get

$$(9b) \qquad |F^{(2)}| < q^{2m} \sum_{n=-\infty}^{\infty} \sum_{\mu=-\infty}^{\infty} [(|n| - \tfrac{1}{2})^2 + (\rho_2 q^2 \mu)^2]^{-m} + q^{2m} \sum_{n=-\infty}^{\infty} (|n| - \tfrac{1}{2})^{-2m} (\rho_2/\rho_1)^{-2m}.$$

Now the first part of the right hand side of (9b), or S_1, can be majorized by an infinite integral if we observe that the summation on μ is of the type $\sum_{-\infty}^{\infty} \phi(\mu)$, where $\phi(\mu)$ decreases monotonically as $|\mu|$ increases. Thus, adding a term to take care of $\phi(0)$, we obtain the estimate

$$S_1 \leq q^{2m} \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} [(|n| - \tfrac{1}{2})^2 + (\rho_2 q^2 \mu)^2]^{-m} d\mu + 2q^{2m} \sum_{n=-\infty}^{\infty} (|n| - \tfrac{1}{2})^{-2m};$$

and, changing the variable of integration, we obtain

$$|F^{(2)}| \leq q^{2m} \rho_2^{-1} q^{-2} \sum_{n=-\infty}^{\infty} \{(|n| - \tfrac{1}{2})^2\}^{-m} d\nu + q^{2m}(2 + (\rho_2/\rho_1)^{-2m}) \cdot \sum_{n=-\infty}^{\infty} (|n| - \tfrac{1}{2})^{-2m}.$$

Since we are in a Stolz neighborhood, as $z - p/q$ $(= \rho_1 + i\rho_2) \to 0$ the ratio $|\rho_2/\rho_1| \geq \lambda$, hence $F^{(2)} = O(\rho_2)^{-1} = O(z - p/q)^{-1}$, whence finally

$$F_{2m}(z) = q^{-2m} 2\zeta(2m)(z - p/q)^{-2m} + O(z - p/q)^{-1}. \tag{9}$$

(The error term will shortly be improved to $o(1)$.)

As in the case of the identically vanishing theta functions, such as $G_4(z)$, the singularities of individual terms do not always reflect the behavior of the series. In general, we shall have to resort to expansion in terms of modular functions to describe behavior at singularities.

7. Use of modular functions: rational singularities. We now consider a theta function $F(z)$ with or without poles and of order m and degree r. Then as $z \to p/q$, a rational number (in lowest terms), in a Stolz neighborhood,

$$f(z) \sim C_0(z - p/q)^{-2m} \exp 2\pi i r q^{-2}(z - p/q)^{-1}.$$

Here the constant C_0 is $k'_0 q^{-2m} \exp - 2\pi i r p' q^{-1}$, where $pp' \equiv 1 \pmod{q}$, and $k'_0 = \lim f(z) \exp 2\pi i r z$ as $\Im z \to \infty$.

The proof of this result is quite simple. We write

$$z' = (p'z - q')/(-qz + p) \text{ where } pp' - qq' = 1.$$

Then $\Im z' \to \infty$ as $z \to p/q$ in a Stolz neighborhood. Now since $f(z') = f(z)(-qz + p)^{2m}$, then $f(z) = q^{-2m})z - p/q)^{-2m} f(z')$. The theorem follows from the asymptotic expression for $f(z')$ combined with $z' = -p'q^{-1} + q^{-1}(-qz + p)^{-1}$.

We observe that the error is $O(\exp[-2\pi i q^{-2}(z - p/q)^{-1}])$, which approaches zero faster than any power of $(z - p/q)$. Thus, for instance, $F_{2m}(z)$, which is of order m and degree 0, can be estimated with principal term as in equation (9) but with error term $o(1)$.

We might add in conclusion that to a certain extent the behavior characterizes the theta function. For, if a theta function satisfies the relation $f(z) \asymp (z - p/q)^s$, for s real, as $z \to p/q$ in a Stolz neighborhood, then $f(z)$ is of degree 0 and $s = -2m$, where m is the order of $f(z)$. This would follow immediately if we knew that $f(z)$ was of finite degree, since then only the degree zero would permit of a behavior which is a power of $(z - p/q)$. To see that $f(z)$ is of finite degree, note that a wide enough Stolz neighborhood about p/q becomes, in the z'-plane, a region wide enough to include the infinite portion of any vertical strip of width 1. Thus, however $\Im z'$ may approach ∞, $f(z')$ is smaller than some power of $(z - p/q)$, which, in turn, is smaller than $\exp - 2\pi i z'$. This proves that $f(z)$ must be of finite degree.

8. Dissection of the neighborhood of an irrational number. The change of variable, which served as a *deus ex machina* in the previous section on rational singularities, can be applied to the irrational singularities with slight modifications.[5]

Our starting point is the simple continued fraction for an irrational number θ. We shall use a few elementary properties of irrational numbers, which can be found, e.g., in Perron [4]. The expansion of θ into a simple continued fraction is written as $\theta = (a_0, a_1, \cdots, a_n, \cdots)$. We denote $(a_0, a_1, \cdots, a_n)$ by p_n/q_n, the convergents. A connection between the denominators of the convergents and the degree of accuracy of the approximation of θ by p_n/q_n is given by

$$a_{n+1}^{-1} < (-)^n q_n^2(\theta - p_n/q_n) < (a_{n+1} + 1)^{-1}, \tag{10a}$$

and the growth of the denominators of the convergents is estimated by

$$a_{n+1} < q_{n+1}/q_n < a_{n+1} + 1. \tag{10b}$$

We are now ready for our major definition: *We say that the complex number z belong to the convergent p_n/q_n of the irrational number θ when* $q_{n+1}^{-2} < \Im z \leq q_n^{-2}$.

THEOREM. *If we let the Stolz neighborhood at θ be of slope $\geq \lambda$, i.e., if we let it lie within the angular region $\Im z \geq \lambda \mid Re\,(z - \theta) \mid$, then under the transformation $z' = (p'z - q')/(-q_n z + p_n)$, $(p_n p' - q_n q' = 1)$, we find that a number z belonging to p_n/q_n with respect to θ goes into a number z' whose imaginary part can be estimated by*

$$\tfrac{1}{2}(a_{n+1} + 1)(2 + \lambda)/\lambda > \Im z' > 2a_{n+1}\,(a_{n+1} + 1)^{-2} \exp - 2/\lambda. \tag{11}$$

Proof. Let us first observe that *the set of points belonging to p_n/q_n and lying in the Stolz neighborhood has a non-euclidean diameter of less than $2/\lambda + 2 \log q_{n+1}/q_n$.* To see this, connect any two points of the region by a path consisting of two segments, one horizontal (of non-euclidean length $< \int dx/y = 2/\lambda$) and one vertical (of non-euclidean length $\leq \int dy/y = 2 \log q_{n+1}/q_n$).

Next perform the transformation $z' = (p'z - q')/(-q_n z + p_n)$. *At*

---

[5] The methods presented in Sections 8 and 9 are a generalization of those of Hardy and Littlewood [3].

least one point belonging to p_n/q_n *must go into a point with* $\Im z' > \frac{1}{2}a_{n+1}$. For, as previously noted, $| \theta - p_n/q_n | < a_{n+1}^{-1}q_n^{-2}$. Thus the Ford circle [1], drawn tangent to the real axis at p_n/q_n, and having radius $a_{n+1}^{-1}q_n^{-2}$, must contain in its interior some points belonging to p_n/q_n, such as $\theta + ia_{n+1}^{-1}q_n^{-2}$ whose imaginary part lies between q_n^{-2} and q_{n+1}^{-2}. But the interior of this Ford circle (which we shall denote, for convenience, by $C_{p_n/q_n}(a_{n+1})$) clearly goes into the portion of the z'-plane with $\Im z' > \frac{1}{2}a_{n+1}$.

We can now prove the second inequality in (11). The presence of a point z'_0 above the line $\Im z' = \frac{1}{2}a_{n+1}$ excludes the presence of a point below the line $\Im z' = h$, if we choose h small enough, (by virtue of the boundedness of the non-euclidean diameter of the set of points belonging to p_n/q_n and hence the boundedness of the z'-image). We need only take h such that $\int_h^{\frac{1}{2}a_{n+1}} dy/y = 2/\lambda + 2 \log q_{n+1}/q_n$. This exceeds the non-euclidean diameter in question. Thus we may take h as $\frac{1}{2}(q_n/q_{n+1})^2 \exp - 2/\lambda$, which leads to the inequality in question when we note that $q_n/q_{n+1} > (a_{n+1} + 1)^{-1}$.

To obtain the first inequality in (11) we first construct a Ford circle about p_n/q_n so small that it excludes all points of the Stolz neighborhood. The circle of radius $\lambda(2+\lambda)^{-1} | \theta - p_n/q_n |$ will suffice for that purpose. For this circle is determined by the condition that its circumscribed square exactly touches one of the lines $y = \lambda | x - \theta |$. This circle clearly contains the smaller circle

$$C_{p_n/q_n}((a_{n+1} + 1)(2 + \lambda)/\lambda), \text{ of radius } \lambda(2+\lambda)^{-1}(a_{n+1} + 1)^{-1}q_n^{-2}.$$

But since no points of the Stolz neighborhood lie inside this circle, it follows that in the z'-plane no images lie on or above the line

$$\Im z' = (2 + \lambda)(a_{n+1} + 1)(2\lambda)^{-1}. \qquad \text{Q. E. D.}$$

We conclude this section by observing that under the previous conditions, z belonging to p_n/q_n and lying in a Stolz neighborhood of θ of slope no less than λ, the quantities $z - \theta$, $-q_n z + p_n$, etc., are interrelated. First of all

$$(a_{n+1} + 1)^{-2}q_n^{-2} < | z - \theta | < (1 + \lambda)q_n^{-2}. \tag{12a}$$

This follows from the relations $\Im z > q_{n+1}^{-2}$, $\Im z \leqq q_n^{-2}$, and $| Re\,(z - \theta)| \leq \lambda q_n^{-2}$. Likewise,

$$\lambda(1+\lambda)^{-1}(a_{n+1} - 1)^{-1}q_n^{-2} < | z - p_n/q_n | < (2 + \lambda)q_n^{-2}. \tag{12b}$$

For the distance from p_n/q_n to the nearest of the two lines $y = \pm \lambda(x - \theta)$ is $\lambda | \theta - p_n/q_n | (1 + \lambda^2)^{-\frac{1}{2}}$, which exceeds $\lambda(1+\lambda)^{-1}q_n^{-2}(a_{n+1} - 1)^{-1}$. On the other hand, the distance $|z - p_n/q_n|$ is less than the sum $|z - \theta| + |\theta - p_n/q_n|$. This justifies the relation (12b). Dividing (12a) by (12b), we obtain the final estimate

$$(12)\quad \lambda^2(1+\lambda)^{-3}(a_{n+1}+1)^{-2} < |-zq_n + p_n|^2/|z-\theta| < (2+\lambda)^2(a_{n+1}+1)^2.$$

9. Behavior of theta functions near irrational singularities. We let θ be an irrational number whose continued fraction has bounded partial quotients. Such numbers θ are known to form a non-enumerable set of measure zero, containing all real quadratic irrationals among others. We were interested in the function $G_{2m}(z)$ since some of the singularities of individual terms of its series occur at real quadratic numbers, but we can just as well carry out the analysis for the larger class of numbers referred to above.

We consider a theta function of finite or infinite degree, without poles, and of order m. Under the change of variable $z' = (p'z - q')/(-q_nz + p_n)$, where $p'p_n - q'q_n = 1$, we obtain: $f(z) = f(z')(-q_nz + p_n)^{-2m}$. If z approaches θ in a Stolz neighborhood of slope less than λ and if θ has bounded partial quotients, then $(-q_nz + p_n)^2 \asymp (z - \theta)$ where z belongs to p_n/q_n. This follows from inequality (12). From inequality (11) we see that $\Im z'$ lies within certain bounds making $f(z') = O(1)$, if z belongs to p_n/q_n. *Thus* $f(z) = O(z-\theta)^{-m}$ *for theta functions of order* m *and without poles.*

If we know that $f(z)$ has *no zeros* as well as no poles, we could write $f(z') \asymp 1$, in the above paragraph, whence $f(z) \asymp (z-\theta)^{-m}$. Applying these results to $G_{2m}(z)$, (theta functions without poles and of order $2m$), we find:

$$(13)\qquad G_{2m}(z) = O(z-\theta)^{-2m},\ m \geq 2,$$

and

$$(13a)\qquad G_{12}(z) \asymp (z-\theta)^{-12},$$

where θ is an irrational number with bounded partial quotients, and $z \to \theta$ in a Stolz neighborhood.

10. Characterization of a real number by the corresponding singularity. We shall see in the concluding paragraphs to what extent the behaviors at real singularities determine the singularity and the function. First we shall see that the behavior near irrationals with bounded partial quotients is seldom better than a *large* O estimation. In fact the estimation (13) can not be improved to the type (13a) unless $m = 6$.

THEOREM. *If θ is an irrational number whose partial quotients a_n do not tend to infinity with n, then a sufficiently wide Stolz neighborhood at θ will include infinitely many points equivalent under the modular group to a preassigned point of the upper half plane.*

(From elementary considerations this infinitude of points would have to converge to θ). For the proof of this theorem, consider the points $A_n = \theta + iq_n^{-2}$, which belong to p_n/q_n. Then if a_{n+1} runs over infinitely many bounded partial quotients, the images A'_n of A_n under the transformation $z' = (p'z - q')/(-q_nz + p_n)$, where $p_np' - q_nq' = 1$, will lie in a horizontal strip $\wp$ by (11), with λ taken as (say) unity. If z^* is the preassigned point of this theorem, let ρ be the smallest radius for which the union of circles about $z^*, z^* \pm 1, z \pm 2, \cdots$ of non-euclidean radius ρ covers the whole strip $\wp$. Then if we draw about the points A_n circles of non-euclidean radius ρ, we should still be able to include these circles in a sufficiently wide Stolz neighborhood. In fact, if we take instead of A_n a continuously moving point $z = \theta + i\eta$, $\eta > 0$, then as η approaches zero continuously, the non-euclidean circles of radius ρ about this moving point sweep out a sector at θ, owing to the invariance of non-euclidean distance under homothetic transformation of the z-plane about θ. Q. E. D.

Thus we can not replace (13) by $G_{2m}(z) \asymp (z - \theta)^{-2m}$ unless $m = 6$, since otherwise, as noted in **5**, $G_{2m}(z)$ has zeros. In fact, *if θ has bounded partial quotients, we can have $f(z) \asymp (z - \theta)^{-2m}$ in any Stolz neighborhood of θ if and only if $f(z)$* is without zeros and poles, which, as noted in **5**, means that *the following holds*:

$$f(z) = G_{12}(z)^{\frac{1}{2}k} \exp g(J(z)), \tag{14}$$

where $g(w)$ is an entire function of w and k is an integer.

We can supplement the last result by noting that a Poincaré theta function $f(z)$ can behave $\asymp (z - \theta)^s$, s real, as $z \to \theta$ in any arbitrary Stolz neighborhood only if θ is rational and $f(z)$ is of degree zero or θ is irrational with bounded partial quotients and $f(z)$ is of the form given in equation (14) above. The proof of this last statement is somewhat tedious, involving the examination of a number of special cases. It will be omitted as it requires no new methods. What we do is let z approach θ vertically and show that if θ does not fall into one of the above categories, then the values taken by the theta function must vary too widely to permit such an asymptotic behavior.

These last results are in contrast with a result of Wintner [7], who showed (essentially) that at almost any point $G_{12}(z)$ has a limit along some path of approach and in fact these paths of approach can even be made to lie in a Stolz neighborhood of preassigned arbitrarily small width. Wintner demonstrated this property for the Dedekind η-function,[6] which has a well-known expansion into powers of $\exp 2\pi i z$, and whose behavior as a function can therefore be analyzed in the light of the behavior of the expansion coefficients.

At any rate it is now apparent that the diophantine nature of a singularity of a theta function governs to a great extent the asymptotic behavior at the singularity, and even the function itself.

HARVARD UNIVERSITY.

BIBLIOGRAPHY.

[1] L. Ford, "A geometrical proof of a theorem of Hurwitz," *Proceedings of the Edinburgh Mathematical Society*, vol. 35 (1917), pp. 59-65.

[2] R. Fueter, *Vorlesungen über die singulären Moduln und die komplexe Multiplikation der elliptischen Funktionen*, vol. 1, Leipzig-Berlin, 1924, p. 29.

[3] G. H. Hardy and J. E. Littlewood, "Some problems in diophantine approximation II," *Acta Mathematica*, vol. 37 (1914), p. 230.

[4] J. F. Koksma, "Diophantische Approximationen," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 4, Berlin, 1936, Chapter IX, 2.

[5] O. Perron, *Die Lehre von den Kettenbrüchen*, Leipzig, 1913, Chapter II.

[6] H. Poincaré, "Fonctions modulaires et fonctions Fuchsiennes," *Œuvres*, vol. II, Paris, 1916, pp. 592-618.

[7] A. Wintner, "A property of the elliptic modular net," *Duke Mathematical Journal*, vol. 12 (1945), pp. 451-454.

---

[6] The proportionality of $G_{12}(z)$ and $\eta(z)^{48}$ was noted in 5 above.

A PROBLEM OF PLANE MEASURE.*

By G. G. LORENTZ.

1. Introduction. Let A be a plane measurable set. For a real x_0 let A_{x_0} be the set of all y, for which the point (x_0, y) belongs to A, that is, the projection upon the y-axis of the section $x = x_0$ of the set A. A_x is measurable for all x except for a null set (that is, a set of zero measure), and we call the function[1]

$$P_A(x) = P(x) = mA_x$$

the cross function of the set A with respect to the x-axis. The set A is not determined by the cross function $P(x)$, but any non-negative measurable function $P(x)$ is a cross function of a measurable set. The chief aim of this paper is to inquire into the conditions under which a *pair* of functions, $P(x)$ and $Q(y)$, is a pair of cross functions of a set A on the x- and y-axes, respectively, and if such is the case, whether A is defined uniquely. We shall consider plane sets A with finite measure $\mu A < +\infty$. Then $P(x)$, $Q(y)$ are almost everywhere finite and integrable over $(-\infty, +\infty)$.

For a non-negative integrable function $P(x)$ defined for $-\infty < x < +\infty$ there exists, according to F. Riesz,[2] a *non-increasing rearrangement* of $P(x)$, that is, a function $p(x)$, $0 < x < +\infty$, equimeasurable with $P(x)$, for which $p(x_1) \geqq p(x_2)$ for $x_1 \leqq x_2$. Here two non-negative integrable functions $p(x)$, $P(x)$ defined on sets e, E respectively, are called *equimeasurable,* if the x-sets defined by $\alpha \leqq p(x)$ and $\alpha \leqq P(x)$ have the same measure for every real α. (Then also other similar sets, e. g. those defined by $p(x) < \alpha$ and $P(x) < \alpha$, have the same measure. In particular, $me = mE$).

Let $y = p(x)$ and $x = q(y)$ be two non-increasing functions defined for $x > 0$, $y > 0$, respectively. We set up the following conditions

$$(1)\qquad \begin{aligned} &\text{(a)}\quad \int_0^x p(u)\,du \leqq \int_0^x q^{-1}(u)\,du, && x > 0,\\ &\text{(b)}\quad \int_0^y q(u)\,du \leqq \int_0^y p^{-1}(u)\,du, && y > 0. \end{aligned}$$

---

* Received April 15, 1948.

[1] In the sequel mB always signifies the *linear* Lebesgue measure of a linear set B, whereas μA signifies the *plane* measure of a set A.

[2] F. Riesz, "Sur une inégalité intégrale," *Journal of the London Mathematical Society,* vol. 5 (1930), pp. 162-168. F. Riesz considers functions on a finite interval.

These conditions are satisfied, for example, and in fact reduce to identities, if $p(x)$ and $q(y)$ are inverses of each other. The functions $p(x) = e^{-x}$, $-2^{-1}\log(y/2)$ satisfy (1) and are not inverses of each other. We can now formulate our main results thus:

THEOREM 1. *Suppose that* $P(x)$, $Q(y)$ *are two non-negative integrable functions defined for* $-\infty < x < +\infty$, $-\infty < y < +\infty$. *In order that a measurable set* A *with cross functions* $P(x)$, $Q(y)$ *exists, it is necessary and sufficient that the non-increasing rearrangements* $p(x)$, $q(y)$ *of these functions satisfy conditions* (1).

THEOREM 2. *The set* A *is determined uniquely modulo null sets by its cross functions if and only if* $p(x)$ *and* $q(y)$ *are inverses of each other.*

In **2** we collect some simple properties of equimeasurable functions and of functions satisfying conditions (1). In **3** we give proofs of Theorems 1 and 2. Finally, in **4** some other problems concerning cross functions are discussed.

2. Lemmas on equimeasurable functions and on functions satisfying conditions (1). In the sequel $p(x)$, $q(y)$ signify non-negative, non-increasing functions defined for $x > 0$ and $y > 0$ respectively, whose integrals over $(0, +\infty)$ are finite. For a function $p(x)$ of this kind, we define $p^{-1}(y)$ to be constant x_0 for $p(x_0-) \leqq y \leqq p(x_0+)$, if x_0 is a point of discontinuity of $p(x)$. On the other hand, if (α, β) is an interval of invariability of $p(x)$ such that $p(x) = y_0$, we choose $p^{-1}(y_0)$ arbitrarily between α and β. Thus

$$\int_0^\infty p(u)\,du = \int_0^\infty p^{-1}(u)\,du. \tag{2}$$

We start with some properties of functions $p(x)$, $q(y)$ satisfying conditions (1). Making $x \to \infty$ and $y \to \infty$ in (1a) and (1b) and using (2) we obtain

$$\int_0^\infty p(x)\,dx = \int_0^\infty q(y)\,dy. \tag{3}$$

Conversely, (1b) follows from (1a) and (3). To prove this, let F be the set of $0 < y < +\infty$ for which the closed intervals $I_1 = [p^{-1}(y+), p^{-1}(y-)]$ and $I_2 = [q(y+), q(y-)]$ have common points. Clearly, the complement of F is open in $y > 0$, and F is closed. Suppose $y \in F$ and let x be a common point of I_1 and I_2. Then

$$\int_0^y p^{-1}(u)\,du + \int_0^x p(u)\,du = \int_0^\infty p(u)\,du + xy,$$

$$\int_0^y q(u)\,du + \int_0^x q^{-1}(u)\,du = \int_0^\infty q(u)\,du + xy.$$

Subtracting we obtain

$$\int_0^y [p^{-1}(u) - q(u)]du = \int_0^x [q^{-1}(u) - p(u)]du.$$

For the left-hand side $\Phi(y)$ of this equality (1a) implies $\Phi(y) \geqq 0$ for $y \varepsilon F$. Now consider a complementary interval $\alpha < u < \beta$ of F. The difference $p^{-1}(u) - q(u)$ is of constant sign there. For otherwise a point $\alpha < y < \beta$ would exist with the property that $p^{-1}(u) - q(u)$ changes sign in every neighborhood of y. Then $p^{-1}(y-) \geqq q(y+)$, $q(y-) \geqq p^{-1}(y+)$, which contradicts the definition of (α, β). Thus $\Phi(y)$ is monotone in every such interval and therefore $\Phi(y) \geqq 0$ there. The same argument applies to the complementary intervals for which $\alpha = 0$ or $\beta = +\infty$, since $\Phi(0+) = \Phi(+\infty) = 0$. This completes the proof.

For equimeasurable functions we need the following lemma:

LEMMA 1. *Suppose that $\Phi(x)$, $x \varepsilon A$ and $\Phi'(x')$, $x' \varepsilon A'$ are equimeasurable functions, and that C is any set of real numbers. If $B \subset A$ and $B' \subset A'$ are sets defined by $\Phi(x) \varepsilon C$ and by $\Phi'(x') \varepsilon C$ respectively, and if B is measurable, then B' is also, and $mB' = mB$.*

Proof. If C is an interval, the assertion follows from the definition of equimeasurable functions. Hence, the same holds for any open set C and any closed set C. Let $B_0 \subset B$ be closed, bounded, and such that $\Phi(x)$ is continuous on B_0. By C_0 we designate the set of all values of $\Phi(x)$ on B_0, and by B'_0 the set of all x' with $\Phi'(x') \varepsilon C_0$. C_0 is closed, hence B'_0 is measurable and $mB'_0 = mB_0$. Now we have (m_*E is the inner Lebesgue measure of a set E)

$$mB = \sup mB_0 = \sup mB'_0 \leqq m_*B'.$$

First suppose $mA = mA' < +\infty$. Similar to the above we have

$$m(A - B) \leqq m_*(A' - B'),$$

and since $mB + m(A - B) = mA' \geqq m_*B' + m_*(A' - B')$,

$$mB = m_*B', \qquad mA - m_*B' = m(A - B) = m_*(A' - B').$$

That is, B' is measurable and $mB' = mB$.

If $mA = +\infty$, we consider instead of A and A' their subsets A_α, A'_α $(\alpha > 0)$ defined by $|\Phi(x)| > \alpha$ and by $|\Phi'(x')| > \alpha$ respectively. They have a finite measure. Thus we get the measurability of $B'A'_\alpha$, and $mBA_\alpha = mB'A'_\alpha$. The same relation for BA_0 and $B'A'_0$ follows by passing to the limit; and thus we obtain the assertion of the lemma, because Φ and Φ' vanish on sets of equal measure.

From this lemma easily follows

LEMMA 2. *If $\Phi(x)$, $\Phi'(x')$ are equimeasurable integrable non-negative functions defined on A and A' respectively, then A, A', except for two null sets, split up in classes of points K, K' such that the corresponding function is constant on every class and that each class is a set of zero measure. Among the classes K and K' a one-to-one correspondence $\mathfrak{F}$ exists such that Φ and Φ' have the same values on classes corresponding to each other. If a set of the form $B = \Sigma K$ is measurable, the corresponding set $B' = \Sigma K'$ is also measurable, and $mB = mB'$. If $\Phi(x)$ is monotonous and A is an interval, then each class K is a single point.*

Proof. The set N of points $x \,\varepsilon\, A$, for which Φ' does not take the value $\Phi(x)$, is a null set according to Lemma 1. We decompose the set $A - N$ in subsets B on which $\Phi(x)$ is constant; they are in one-to-one correspondence with the analogous subsets B' of $A' - N'$ (N' is defined similarly to N). These subsets B, B' are classes K, K' in the case they are null sets. If, on the other hand, $mB = mB' > 0$ (these measures are necessarily finite), we define on B and B' the functions $\phi(x) = m[B \cdot (-\infty, x)]$, $\phi'(x') = m[B' \cdot (-\infty, x')]$. Then the classes K, K' are the subsets of B and B' on which ϕ and ϕ' take the same values. It is easily seen by Lemma 1 that the classes K, K' and the correspondence $\mathfrak{F}$ thus defined have all the announced properties.

3. Proof of Theorems 1 and 2. 3.1. First of all we shall show: If $P(x)$, $Q(y)$ are the cross functions of a measurable set A with $\mu A < +\infty$, and if $p(x)$, $q(y)$ are their non-increasing rearrangements, then p, q satisfy conditions (1). (This is the first part of Theorem 1).

Let A_1 be the set obtained by "pouring" A on the x-axis, that is the "ordinate set" of $P(x)$, defined by $-\infty < x < +\infty$, $0 \leqq y \leqq P(x)$. A_1 has the cross functions $P_1(x) = P(x)$ and $Q_1(y)$, and $Q_1(y)$ is non-increasing. We assert:

$$\int_B Q(u)\,du \leq \int_0^v Q_1(u)\,du \tag{4}$$

for every $v \geqq 0$, if $mB \leqq v$. In fact, let C be the set of points $(x, y) \,\varepsilon\, A$ with $y \,\varepsilon\, B$ and C_1 be the set of points $(x, y) \,\varepsilon\, A_1$ with $0 \leqq y \leqq v$. Then $P_C(x) \leqq P(x)$ and, further, $P_C(x) \leqq v$. Therefore $P_C(x) \leqq \operatorname{Min}[P(x), v] = P_{C_1}(x)$. An integration over $(-\infty, +\infty)$ gives

$$\mu C \leqq \mu C_1. \tag{5}$$

Thus (4) is established. From (4) it follows immediately for the rearrangement $q(y)$ of $Q(y)$

$$\int_0^y q(u)\,du \leqq \int_0^y Q_1(u)\,du, \qquad y \geqq 0. \tag{6}$$

Now we pour A_1 on the y-axis, and get a set A_2 with the cross functions $P_2(x)$ and $Q_1(y)$, which are both non-increasing. A_2 is bounded by the axes and by the curve $y = P_2(x) = Q_1^{-1}(x)$. Applying (6) to this second "pouring" we get

$$\int_0^x p(u)\,du \leqq \int_0^x P_2(u)\,du = \int_0^x Q_1^{-1}(u)\,du, \qquad x \geqq 0.$$

According to 2 and by making use of $\mu A_1 = \mu A_2$ this becomes

$$\int_0^y Q_1(u)\,du \leqq \int_0^y p^{-1}(u)\,du. \tag{7}$$

(6) and (7) imply (1b), and in a similar way (1a) may be proved.

3.2. Suppose that $p(x)$, $q(y)$ are functions satisfying conditions (1). We shall show that a set A with the cross functions p, q exists, and further that there are two such sets, essentially different from one another (that is, not equivalent modulo null sets), when the relation $p = q^{-1}$ is not fulfilled.

Let D be the set of all points (x, y), $x > 0$, $y > 0$ which satisfy $x \leqq q(y)$ and $x \leqq p^{-1}(y)$; and let D^+ and D^- be defined by $q(y) < x < p^{-1}(y)$ and by $p^{-1}(y) < x < q(y)$ respectively (we only take into consideration points of continuity y of p^{-1} and q). The sets D^+, D^- are open; their projections on the y- and the x-axis are disjoint open sets. For any plane set A we denote by $A(y_0)$ the part of A which is situated in the strip $0 < y < y_0$. By (1b) we have then

$$\mu D^-(y) \leqq \mu D^+(y), \qquad y \geqq 0. \tag{8}$$

We wish to establish the representations

$$D^+ = \Sigma R_i^+ + N^+, \qquad D^- = \Sigma R_i^- + N^-, \tag{9}$$

where N^+, N^- are null sets and R_i^+, R_i^- are open "net-squares" of the kind

$$2^{-n}l < x < 2^{-n}(l+1), \qquad 2^{-n}m < y < 2^{-n}(m+1), \qquad l, m, n = 0, 1, 2, \cdots$$

such that R^+_i and R_i^- are congruent, and that R_i^- lies above and on the left of R_i^+.

For we can represent D^- in the form $\Sigma R_i + N$, where R_i are open net squares. The square R_1 lies in a strip $\alpha < y < \beta$, which contains no points of D^+. Since, by (8), $\mu D^+(\alpha) \geqq \mu D^-(\beta)$, there is a γ, $0 \leqq \gamma < \alpha$ for which $\mu[D^+(\alpha) - D^+(\gamma)] = \mu R_1$. The open set $D^+(\alpha) - D^+(\gamma)$ lies below and on the right of R_1; thus, representations (9) exist for $D^+(\alpha) - D^+(\gamma)$ and R_1. The same process, applied to $R_2, R_3, \cdots$ establishes the result. We shall assume, as we may, that in (9) the squares are arranged according to their size.

By means of (9) we prove assertion 2 in the following way. The set $A_0 = D + D^+$ has the cross functions $p(x)$ and $p^{-1}(y)$. We obtain the set A of the assertion by shifting certain squares of A_0 vertically upwards on to an empty place. Thus the new y-cross function becomes the same as for $D + D^+$, that is $q(y)$, while the x-cross function remains unaltered.

These shiftings are defined inductively. Suppose that the $1, 2, \cdots$, $(k-1)$-st steps have been taken, leading to the set A_{k-1}. Further let $\alpha < y < \beta$ be the strip in which R_k^+ lies and $\gamma < y < \delta$, $(\gamma \geqq \beta)$ the strip of R_k^-. Put $l = \beta - \alpha = \delta - \gamma$. Then

$$Q_{A_{k-1}}(\beta) \geqq q(\beta) + l \geqq q(\gamma) + l,$$
$$Q_{A_{k-1}}(\gamma) \leqq q(\gamma) - l.$$

The section $(A_{k-1})_\beta$ of A_{k-1} contains at least $[q(\gamma)/l] + 1$ net-intervals of length l (because $(A_0)_\beta$ was a single interval from which only net intervals of lengths $\geqq l$ were removed); in the same way the section $(A_{k-1})_\gamma$ is contained in $[q(\gamma)/l] + 1 - 1 < [q(\gamma)/l] + 1$ such intervals. Therefore, there is in the strip $\alpha < y < \beta$ a full square R of A_{k-1}, over which an empty place in the strip $\gamma < y < \delta$ exists. The k-th step consists in the vertical displacement of R into the strip $\gamma < y < \delta$. This displacement thus carries R from a strip containing points of D^+ (and no points of D^-) into one containing points of D^- (and none of D^+).

The set A thus formed has the required cross function,

$$\begin{aligned} Q_A(y) &= Q_{A_0}(y) - \Sigma Q_{R_i^+}(y) + \Sigma Q_{R_i^-}(y) \\ &= Q_{A_0}(y) - Q_{D^+}(y) + Q_{D^-}(y) = Q_{D+D^-}(y) = q(y). \end{aligned}$$

If $p(x) = q^{-1}(x)$ does not hold for all x, then there are two different sets A, A' with the cross functions $p(x)$, $q(y)$. In fact, at the first step of the above construction, we could shift either R_1^+ itself upwards or the square of the same horizontal strip lying exactly under R_1^-. Consequently there are sets A with cross functions p, q *containing no points of* R_1^+ and such containing R_1^-. Interchanging the role of x and y, that is of p and q, and of R_1^+ and R_1^-, we see that there exists also such a set A *containing* R_1^+ Consequently there are at least two different sets A.

3. 3. The set A with the cross functions p, q constructed in **3. 2** clearly has the form

$$A = \sum_{i=1}^{\infty} E_i + N - \left(\sum_{i=1}^{\infty} \bar{E}_i + \bar{N} \right) \tag{10}$$

with null sets N, $\bar{N}$ and "elementary sets" E_i, $\bar{E}_i$. (By a plane elementary set we understand the product of two measurable linear sets; in the construction above these E_i, $\bar{E}_i$ were squares.)

Let A be a set of the form (10) with cross functions $p(x)$ and $Q(y)$ ($p(x)$ being non-increasing); if moreover $P(x)$ is equimeasurable with $p(x)$, then there is a set of the form (10) with cross functions $P(x)$ and $Q(y)$. In fact, suppose that B, B' are sets on which p, P respectively do not vanish. According to Lemma 2 of **2**, a class K' of points x' of B' corresponds to almost every $x \varepsilon B$. We project the set A_x on the vertical lines through every such $x' \varepsilon K'$. This process, which may be designated by $\mathfrak{P}$, leads to the required set A'. For A' has the cross functions $P(x)$ and $Q(y)$. Finally, A' is measurable, being again of the form (10) (for $\mathfrak{P}$ preserves elementary sets).

The second part of Theorem 1 is now obtained by applying $\mathfrak{P}$ twice.

Finally the necessity of the condition $p(x) = q^{-1}(x)$ in Theorem 2 results. For if this condition is not fulfilled, there are two essentially different sets A, $\bar{A}$ with cross functions p, q. Then $P_{A-\bar{A}}(x) > 0$ on a set of positive measure. Since this latter condition is preserved under $\mathfrak{P}$, the above reasoning shows that there are two essentially different sets with cross functions P and Q.

3. 4. In order to prove Theorem 2 it remains to show that, if $p(x) = q^{-1}(x)$, only one solution of the problem exists. We first observe that the set A, constructed in **3. 2** and **3. 3**, has the following property:

(11) *For almost all* y_1, y_2 *the inequality* $Q(y_1) \leqq Q(y_2)$ *implies* $A_{y_1} \subset A_{y_2}$.
For (11) holds for the set A bounded by the curve $y = p(x)$ and by the axes, with which we started the construction; and (11) is preserved under $\mathfrak{P}$.

Suppose now that A is a set with cross functions P, Q, and that A has the property (11). We shall show that no set A' essentially different from A and with the same cross functions can exist. For such a set the measure of $A - A'$ would be positive. Therefore, we can choose an $\epsilon > 0$ and a y-set B of positive measure so that

$$m(A_y - A'_y) \geqq \epsilon, \qquad y \varepsilon B. \tag{13}$$

We may assume that $Q(y)$ is continuous on B, and we choose a point y_0 of density 1 of this set. Let C be the set of (x, y) defined by $x \varepsilon A_{y_0}$, $-\infty < y < +\infty$. Then

$$\mu CA = \int_{A_{y_0}} p(x)\,dx = \mu CA'$$

and therefore

$$\int_{-\infty}^{+\infty} mCA_y dy = \int_{-\infty}^{+\infty} mCA'_y dy. \tag{13}$$

If $Q(y) \leqq Q(y_0)$, we have $A_y \subset C$ in accordance with (11) and therefore

$$mCA_y = Q(y) = mA'_y \geqq mCA'_y;$$

and if $Q(y) \geqq Q(y_0)$,

$$mCA_y = mC \geqq mCA'_y.$$

Hence $mCA'_y \leqq mCA_y$ almost everywhere, and (13) implies $mCA'_y = mCA_y$ almost everywhere. We shall show that this contradicts (12). Let $\epsilon > 0$ be arbitrary; we choose $\delta > 0$ such that

$$|\, mA_y - mA_{y_0} \,| < \epsilon/2, \qquad\qquad y \varepsilon B,\ |\, y - y_0 \,| < \delta.$$

Then $mCA_y > mA_{y_0} - \epsilon/2$ for these y; moreover $mD_y < mA_{y_0} + \epsilon/2$, if the set D is defined by $x \varepsilon A_y + A_{y_0}$, y arbitrary. On the other hand, because of $A_y \subset D$ and (12),

$$m(D - A'_y) = m(D - DA'_y) \geqq \epsilon;$$

hence

$$mCA'_y \leqq mDA'_y \leqq mD - \epsilon < mA_{y_0} - \epsilon/2 < mCA_y$$

for all y from a certain subset of B of positive measure, in contradiction to what was proved above.

4. Similar problems. 4.1. Up to now we have assumed that the cross functions are integrable, that is, that the set A has a finite measure. But the problems treated have also a sense in the case $\mu A = +\infty$. However, they cannot be dealt with by the method applied before, because a non-integrable function $P(x)$ does not generally possess a monotone rearrangement. We shall deal with a special case to show that matters are now quite different than if $\mu A < +\infty$.

THEOREM 3. *Suppose that $P(x)$, $Q(y)$ are two continuous increasing functions, defined for $x \geqq 0$, $y \geqq 0$ respectively, and vanishing at the origin.*

Then there exists exactly one set A, bounded by two curves of the form $y=f(x)$, $y=g(x)$, $f(x)\leqq g(x)$ with continuous increasing functions $f(x)$, $g(x)$, vanishing for $x=0$, and with cross functions P, Q.

In other words there is exactly one pair of functions f, g of the kind mentioned satisfying the system of functional equations

$$\left\{\begin{aligned} g(x)-f(x)&=P(x),\\ f^{-1}(y)-g^{-1}(y)&=Q(y).\end{aligned}\right. \tag{14}$$

Proof. We put

$$f_0(x)=0, \qquad g_0(x)=P(x) \tag{15}$$

and

$$f^{-1}{}_n(y)=g^{-1}{}_{n-1}(y)+Q(y), \quad g_n(x)=f_n(x)+P(x), \quad n=1,2,\cdots. \tag{16}$$

Except for f_0, all these functions are increasing and continuous; they vanish at the origin. From (16) follows easily: if $f_n(x)\geqq f_{n-1}(x)$ for all $x\geqq 0$, then also $f_{n+1}(x)\geqq f_n(x)$. Therefore the limiting functions $f(x)=\lim f_n(x)$, $g(x)=\lim g_n(x)$ exist. They are finite, since $g_n(x)\leqq P(x)$, and continuous, on account of $g_n(x')-g_n(x)\leqq P(x')-P(x)$ for $x\leqq x'$. Furthermore they are monotone increasing. Making $n\to\infty$ in (16) we see that f, g satisfy equations (14).

Finally f, g is the only solution. For let f', g' be another solution and f not identical with f', say $f'(a)-f(a)>0$ for an $a>0$. Let x_0 be the point of $[0,a]$ where $f'(x)-f(x)$ reaches its maximum. We put $y_0=f(x_0)$, $y'_0=f'(x_0)$, further

$$\begin{aligned} x_1&=g^{-1}(y_0)=f^{-1}(y_0)-Q(y_0)=x_0-Q(y_0),\\ x'_1&=g'^{-1}(y_0)=x_0-Q(y'_0),\end{aligned}$$

then $0<x'_1<x_1<x_0$. Hence

$$\begin{aligned} f(x_1)&=g(x_1)-P(x_1)=y_0-P(x_1),\\ f'(x'_1)&=y'_0-P(x'_1)>y'_0-P(x_1);\end{aligned}$$

and therefore

$$f'(x_1)-f(x_1)>f'(x_0)-f(x_0),$$

contradicting the definition of x_0.

4. 2. We may ask which properties the functions $P(x)$, $Q(y)$ must have in order that a corresponding convex set A exists. A necessary condition is that the functions $P(x)$, $Q(y)$ be *convex*. But this is not sufficient, as is shown by the following example. Let

$$P(x) = Q(x) = \begin{cases} x/2 & \text{for } 0 \leqq x \leqq 1, \\ (2-x)/2 & \text{for } 1 \leqq x \leqq 2, \\ 0 \text{ otherwise.} \end{cases}$$

If a convex region A existed for P, Q, it would have boundary points on each side of the square $0 \leqq x \leqq 2$, $0 \leqq y \leqq 2$. But these boundary points can only lie at the edges of the square. For if, e. g., $(0, y_0)$, $0 < y_0 < 2$, were a boundary point of A, then A, at least in the neighbourhood of $(0, y_0)$, would contain the interior of the angle $(2,0)(0,y_0)(2,2)$, and this interior has an x-cross function $> x > P(x)$. Because of the symmetry we may assume $(0,0) \, \varepsilon \, A$, then also $(2,2) \, \varepsilon \, A$. So A is bounded by two curves $y = f(x)$, $y = g(x)$ ($f(x) \leqq g(x)$, f, g monotone) joining $(0,0)$ and $(2,2)$. By conclusions similar to those in **4. 1** we can show that exactly one pair of monotone functions f, g exists for $0 \leqq x \leqq 1$ and $0 \leqq y \leqq 1$, satisfying conditions (14). Such a pair is $f(x) = (17^{\frac{1}{2}} - 1)x/4$, $g(x) = (17^{\frac{1}{2}} + 1)x/4$. By continuation, this is still true in some intervals $0 \leqq x \leqq \alpha$, $0 \leqq y \leqq \alpha$, $\alpha > 1$. Considering in the same way the course of the curves in the neighbourhood of $(2,2)$ we find $2 - g(x) = (17^{\frac{1}{2}} - 1)(2-x)/4$ in the interval $1 \leqq x \leqq 2$, contradicting the preceding value of $g(x)$ for $1 < x < \alpha$.

4. 3. The cross function of a plane set A with respect to a direction x is the cross function of A with respect to any axis of this direction. A set A is not always uniquely determined by its cross functions with respect to two directions. The same holds for any finite system $x_1, x_2, \cdots, x_n$ of directions: *Two disjoint bounded sets A, B exist possessing the same cross functions with respect to these directions.* For $n = 1$ the assertion is trivial, and in the general case it is proved by induction. Suppose that A_{n-1}, B_{n-1} are two disjoint bounded sets with equal cross functions with respect to $x_1, \cdots, x_{n-1}$. We displace $A_{n-1} + B_{n-1}$ parallel to itself in the direction perpendicular to x_n so that the new set $A'_{n-1} + B'_{n-1}$ has no points in common with $A_{n-1} + B_{n-1}$. Then $A_n = A_{n-1} + B'_{n-1}$ and $B_n = A'_{n-1} + B_{n-1}$ are disjoint and have equal cross functions with respect to $x_1, \cdots, x_n$.

Further possible questions are these: Is a measurable[3] set uniquely determined by its cross functions with respect to *all* straight lines? Is a set necessarily convex if all its cross functions are convex?

TÜBINGEN, GERMANY.

---

[3] For non-measurable sets the answer is negative. There are non-measurable sets for which all cross functions vanish (W. Sierpinski, *Fundamenta Mathematicae*, vol. 1 (1920), pp. 112-115.

DIFFERENTIAL GEOMETRY OF A GENERAL MOTION IN THE COMPLEX PROJECTIVE LINE.*

By P. O. Bell.

---

1. Introduction. The purpose of the author in the present paper is to study the complex projective differential geometry of a general motion of a point in the complex line. Such a motion may be conveniently portrayed as a motion of a point along a curve in the Cauchy plane. The fundamental group of this geometry is the group of complex projective transformations of the points of the complex line and the methods are the usual methods of projective differential geometry. The methods employed in parts of **2** and **3** are similar to those which E. Cartan [1, pp. 28-42] has used in his study of the geometry of the motions of a point in the complex line subject to the group of real projective transformations. The moving reference frame employed by Cartan owes its geometric significance to the preservation of the Poincaré half-plane [1, pp. 33-34] by the real projective transformations of the points of the complex line. The present study calls for a local reference frame (for the definition of projective homogeneous coordinates) which possesses geometric significance with respect to the larger group of complex projective transformations. Such a reference frame attached to a moving point of a general curve of the Cauchy plane is introduced and geometrically characterized in the present paper.

Throughout the present paper the point of view adopted is that of kinematics. The motion of a point $z(t)$ will be classified according to its relation to time t as independent parameter. According as the real and imaginary parts of the Schwarzian derivative $\{z, t\}$ satisfy one of the following conditions:

$$(1)\quad I \neq 0, \qquad (2)\quad I = 0, R < 0, \qquad (3)\quad I = 0, R > 0, \text{ or } (4)\quad I = R = 0,$$

the motion of z at an instant t will be called (1) loxodromic, (2) hyperbolic, (3) elliptic, or (4) parabolic, respectively. Each type of motion is associated with a homography of a type referred to by the same name as follows: A curve C_w is uniquely determined by the following properties: it is tangent to C at $z(t_0)$, and the non homogeneous coordinate w of its generic point

---

* Received March 30, 1948; revised January 6, 1949.

$w(t)$ satisfies the condition $\{w, t\} = \{z, t\}_{t_0}$. A displacement of points of C_w produced by an arbitrary parametric translation $t = t' + h$ can be produced by a homography of the type named. These four categories of homographies are entirely equivalent to the classical ones which have the same names (cf. Osgood [2, pp. 262-270]). The kinematic characterizations of the types (2), (3) and (4) reduce to those of Cartan [1, pp. 39-42] upon restricting the group to be that of the real projective transformations of a point in the complex line. The curves of most general type which can be mapped upon themselves by a transformation (other than the identity) of the group are loxodromes. If the fixed points of the mapping consist of a finite point and the infinite point, the loxodrome becomes a logarithmic spiral. Circles appear as special loxodromes which correspond to real values of a certain invariant function. Since the loxodromes appear as basic curves in the classical theory of the loxodromic homographies when viewed from the standpoint of the theory of functions of a complex variable [2, pp. 261-262] it is not surprising that they play an important role in the present theory.

Various authors have employed the invariants known as inversive arc length and inversive curvature in their studies of inverse differential geometry of plane curves, notably Mullins (in a Columbia dissertation in 1917), H. Liebmann [3], T. Kubota [4], J. Maeda [5], F. Morley [6], and B. C. Patterson [7]. Inversive arc length of a curve C is a real parameter p such that at a generic point $z(p)$ of the curve the imaginary coefficient of the Schwarzian derivative of z with respect to p is constantly equal to 1. The real part of this Schwarzian derivative is the differential invariant known as inversive curvature. These invariants would serve as a basis for the study of the intrinsic differential geometry of plane curves under homographies, since the group of homographies is a subgroup of the inversive group. However, since the inversive arc length vanishes at all points of a circle and at points of a curve where the ordinary curvature is constant, it is highly desirable for the purposes of a kinematic study to introduce a metric so characterized that except at a parabolic point of the curve the associated arc-length does not vanish along the curve. Such an arc length called *the homographic arc length* of a curve C will be defined in terms of an absolute (ζ, η) geometrically determined at each non-parabolic point of the curve. The element of homographic arc length also possesses advantages of symmetry and in addition actually characterizes a euclidean geometry with respect to homographies. Along a "non-euclidean straight line" in the half-plane of Poincaré [8, pp. 7-8] on which the Poincaré metric is defined, the element of homographic arc length becomes identical with the element defined by the

Poincaré metric. The geometric characterization of the element of homographic are length of C at $z(t)$ will be expressed very simply in terms of a certain angle θ and its differential $d\theta$ whose determinations depend upon the particular type of motion of the point $z(t)$ at the instant t.

The *homographic distance* between two points P, Q with respect to an absolute (ζ, η) is defined by

$$\overline{PQ} = |\log(\zeta\eta PQ)|,$$

in which the right member is the absolute value of the logarithm of the cross-ratio of the four complex points ζ, η, P, Q. A "straight line" with respect to this metric (which will be called an s-line) will be defined to be a curve whose homographic arc length between any two of its points is equal to the homographic distance between these two points. An s-line is found to be a curve which can be mapped upon itself by a homography whose double points are ζ and η; the general s-line is a loxodrome, special cases are logarithmic spirals, arcs of circles having ζ and η as endpoints, and the circles which are orthogonal trajectories of these arcs. The homographic metric defined with respect to a given absolute ζ, η and the plane geometry of this metric are found to be *euclidean*; a simple conformal mapping of the s-lines of the plane upon the ordinary euclidean straight lines of the plane is obtained.

2. Normal coordinates and the Schwarzian derivative. Let C denote a curve described in the Cauchy plane by a moving point z whose homogeneous complex projective coordinates are taken to be $(1, z(t))$ where t is a real parameter. The coordinates of z and those of any projective transform of z satisfy a second order differential equation of the form

$$z'' + pz' = 0 \tag{2.1}$$

in which p is a function of t and the accent denotes differentiation with respect to t. Curves projectively equivalent to C as well as the curve C are, therefore, integral curves of (2.1). To normalize homogeneous coordinates of the point z let a proportionality factor λ be introduced by the formula $z = \lambda\omega$. Equation (2.1) assumes the form

$$\lambda\omega'' + (2\lambda' + p\lambda)\omega' + (\lambda'' + p\lambda')\omega = 0.$$

Let λ be chosen so that the coefficient of ω' vanishes, that is $\lambda = \exp(-\frac{1}{2}\int p\,dt)$. With this choice of λ the differential equation of the curve reduces to the form

$$\omega'' + r\omega = 0 \tag{2.2}$$

where r is an invariant defined by $r = -(2p' + p^2)/4$. It may also be shown that r may be expressed in terms of the non-homogeneous coordinate z by the relations

$$2r = \{z, t\} = z'''/z' - \tfrac{3}{2}(z''/z')^2$$

that is, $2r$ is equal to the Schwarzian derivative of z with respect to t. Moreover, the proportionality factor λ can be expressed by the formula $\lambda = (z')^{-\frac{1}{2}}$; so that the normal projective homogeneous coordinates of the point z are given by the relations $\omega^1 = 1/\sqrt{z'}$, $\omega^2 = z/\sqrt{z'}$.

The author has shown in a former paper [9, p. 490] that the Schwarzian derivative can be expressed in terms of euclidean curvature γ and euclidean arc length s by means of the formula

$$\{z, t\} \equiv (id\gamma/ds + \gamma^2/2)(ds/dt)^2 + \{s, t\}. \tag{2.3}$$

On making use of the formula for interchanging the variables in the Schwarzian derivative

$$\{s, t\}dt^2 + \{t, s\}ds^2 = 0,$$

equation (2. 3) may be made to assume the form

$$\{z, t\} \equiv (id\gamma/ds + \gamma^2/2 - \{t, s\})(ds/dt)^2. \tag{2.4}$$

The real and imaginary components of r,

$$R = (\gamma^2 - 2\{t, s\})(ds/dt)^2/4, \qquad I = (d\gamma/ds)(ds/dt)^2/2$$

are, themselves, absolute invariants of the complex projective transformations

$$\rho\Omega^1 = c\omega^2 + d\omega^1, \qquad \rho\Omega^2 = a\omega^2 + b\omega^1, \tag{2.5}$$

in which the coefficients are complex constants, the non-homogeneous coordinates of z and Z being defined by $z = \omega^2/\omega^1$ and $Z = \Omega^2/\Omega^1$, respectively. The following are, therefore, invariant equations

$$(\gamma^2/2 - \{t, s\})ds^2 = (\bar{\gamma}^2/2 - \{t, \bar{s}\})d\bar{s}^2, \qquad d\gamma ds = d\bar{\gamma}d\bar{s}$$

where $\gamma, \bar{\gamma}$ and $ds, d\bar{s}$ are corresponding euclidean curvatures and arc length elements of a curve C and its transform $\bar{C}$ by a transformation (2. 5). The transformations (2. 5) may be written in the form of the homographies

$$Z = (az + b)/(cz + d) \tag{2.6}$$

in which a, b, c, d are complex constants.

3. Homographic mappings of curves upon themselves; fixed points of the mappings. A parametric mapping of a curve C upon itself is determined by two relations of the form

$$\Omega(t)=\omega(\bar{t}), \qquad \bar{t}=\bar{t}(t) \tag{3.1}$$

where $\bar{t}$ is a single valued function of t. The curve C may be denoted by C_t or $C_{\bar{t}}$ according as $\omega(t)$ or $\omega(\bar{t})$ is regarded as the generic point, respectively. Certain curves exist each of which possesses the property that a parametric mapping function $\bar{t}(t)$ exists which maps an arbitrarily chosen point ω_0 of the curve upon another arbitrarily chosen point Ω_0 of the curve and simultaneously maps the curve upon itself in such a manner that the mapping can be obtained by a homography

$$z(\bar{t})=(az(t)+b)/(cz(t)+d) \tag{3.2}$$

where $z=\omega^2/\omega^1$. These curves will now be determined.

The equations

$$d^2\omega(\bar{t})/d\bar{t}^2+r(\bar{t})\omega(\bar{t})=0, \qquad d^2\Omega(t)/dt^2+r(t)\Omega(t)=0, \tag{3.3}$$

are satisfied identically in $\bar{t}$ and t, respectively. The first of these is merely the differential equation of C, and the second expresses the condition that the curve $C_{\bar{t}}$ be equivalent to C_t with respect to a transformation (3.2). On substituting for $d^2\Omega/dt^2$ in (3.3) the right member of the relation

$$d^2\Omega/dt^2=(d^2\omega/d\bar{t}^2)(d\bar{t}/dt)^2+(d\omega/d\bar{t})(d^2\bar{t}/dt^2)$$

one obtains the equation

$$d^2\omega(\bar{t})/d\bar{t}^2+(d\omega/d\bar{t})(d^2\bar{t}/dt^2)(dt/d\bar{t})^2+r(t)\omega(\bar{t})(dt/d\bar{t})^2=0.$$

Comparison of this equation with the first equation of (3.3) yields the relations

$$d^2\bar{t}/dt^2=0, \qquad r(\bar{t})=r(t)(dt/d\bar{t})^2, \tag{3.4}$$

because the points $\omega(\bar{t})$, $d\omega/d\bar{t}$, $d^2\omega/d\bar{t}^2$ are assumed to be linearly independent. From the first of these relations it follows that the mapping functions $\bar{t}$ are of the form

$$\bar{t}=kt+h, \qquad k,h=\text{const.} \tag{3.4}$$

The second relation therefore becomes

$$k^2 r(kt + h) = r(t). \tag{3.5}$$

The constant h may be arbitrarily selected, since no further restrictions are imposed upon $\bar{t}$ than those given by equations (3.4). To determine k, hold t and k fixed and allow h to vary. It follows from (3.5) that $r(kt + h)$ is a constant equal to $r(t)$ and $k = \pm 1$. *The curves whose parametric mappings upon themselves can be obtained by homographies are therefore the* $r(t) = $ *constant curves, and the mapping functions are of the forms* $\bar{t} = \pm t + h$. These mapping functions serve to define what will be called *direct* and *indirect parametric displacements* of the points of a curve C according as the sense of variation of the parameter is maintained or is reversed, respectively. It is clear now that an arbitrary pair of points ω_0, Ω_0 may be put in correspondence by a direct (or an indirect) displacement with a suitable choice of the constant h.

Let $r = R + iI$. A transformation which maps an $r = $ const. curve upon itself will be called (i) loxodromic, (ii) hyperbolic, (iii) elliptic, or (iv) parabolic, according as (i) $I \neq 0$, (ii) $I = 0, R < 0$, (iii) $I = 0, R > 0$, or (iv) $I = R = 0$, respectively. If t is taken to be time, a kinematic interpretation of the description of a curve by a generic point $\omega(t)$ is obtained upon considering the parametric displacements of type (i) to (iv). A motion along an $r = $ const. curve during a lapse of time may therefore be called loxodromic, hyperbolic, elliptic, or parabolic, according to the stipulations upon r stated above.

Since the imaginary component of r is given by $I = (d\gamma/ds)(ds/dt)^2/2$ and $ds/dt \neq 0$, it follows that $d\gamma/ds = 0$ if $I = 0$. Hence, *the* $I(r) = 0$ *curves are circles.*

The most general $r = $ const. curves are integral curves of a differential equation of form (1.2). The homogeneous coordinates of a general point ω of such a curve are therefore given by the form

$$\omega = c_1 e^{-i\sqrt{r}t} + c_2 e^{i\sqrt{r}t}. \tag{3.6}$$

These curves are, clearly, projectively equivalent to the curve defined by $\omega^1 = 1$, $\omega^2 = e^{2i\sqrt{r}t}$, that is, the curve defined in a non-homogeneous coordinate by

$$z = e^{2i\sqrt{r}t}. \tag{3.7}$$

It follows *that the* $r = $ *real const. curves are circles.*

If $r = \text{const.}$, but $I \neq 0$, it follows that the two square roots of r may be written as $\sqrt{r}$, $-\sqrt{r}$ where $\sqrt{r} = \mu + i\nu$, $\nu \neq 0$. Equation (3.7) may be written in the form

$$(3.8) \qquad \log z = -2\nu t + 2i\mu t.$$

Upon substituting $\theta/2\mu$ for t this equation assumes the form

$$(3.9) \qquad \log z = -\nu\theta/\mu + i\theta, \qquad \nu/\mu = \text{const.}$$

The curve represented by this equation is a logarithmic spiral, and the curve (3.6) *of most general type which is projectively equivalent to it is a non-degenerate loxodrome.*

An $I = 0$ curve may be mapped upon itself by a transformation of one of the last three types, the type being determined by the nature of the parameter used in the displacement.

THEOREM. *A curve obtained by transforming a given non-degenerate logarithmic spiral by a homography* (2.6) *in which either c or d is equal to zero is a logarithmic spiral which can be superposed upon the original curve by means of a euclidean rigid displacement.*

To prove this theorem it is sufficient to determine the effects of (1) an expansion and rotation $Z = az$, and (2) an inversion $Z = 1/z$. A translation, of course, merely displaces the curve. As z describes a given logarithmic spiral defined by

$$(3.10) \qquad \log z = k\theta + i\theta \qquad k = \text{const.} \neq 0,$$

the locus of a point Z defined by $Z = az$, where $\log a = \alpha + i\beta$, is determined by the equation $\log Z = k\theta + \alpha + i(\theta + \beta)$. This may be written in the form $\log Z = k\bar{\theta} + i\bar{\theta} + i(\beta - \alpha/k)$, where $\bar{\theta} = \theta + \alpha/k$. The locus thus defined is therefore that of the given logarithmic spiral rotated through the angle $\beta - \alpha/k$. Hence the transformed spiral can be superposed upon the original spiral by a rigid euclidean displacement. As z describes the locus (3.10) the point Z given by $Z = 1/z$ moves along the same curve, since the locus of Z is represented by the equation

$$(3.11) \qquad \log Z = -\log z = -k\theta - i\theta,$$

the angle of Z being the negative of the angle of z. This completes the proof.

Let ψ denote the angle which the tangent to a logarithmic spiral $r = \text{const.}$

makes with the radius vector ρ drawn from the pole (asymptotic point of the logarithmic spiral) to the point of contact of the tangent. The tangent of this angle is defined by the relation $\tan\psi = \rho d\theta/d\rho = -\mu/\nu$. Since $\sqrt{r}$ is given by $\mu + i\nu$ and it is invariant of the transformations (3.6), $\tan\psi$ is itself an invariant. Let ϕ denote the angle of $\mu + i\nu$. It follows that $\tan\phi\tan\psi = -1$, and therefore, one of the relations $\psi = \phi \pm \pi/2$ is satisfied. If the direction of the $\theta =$ const. curve is known at a point, the direction of the point of the loxodrome which corresponds to a given value of r is easily determined by this relation.

To determine the fixed points of a homographic mapping which maps a curve C upon itself, let $\omega(t)$ be a generic point of the curve. Then any point P of the plane is given by an expression of the form $\omega + u\omega'$ where u is a suitable complex number. Let u be a constant to be determined by the condition that the point P be a fixed point. This condition is represented by a relation of the form $(\omega + u\omega')' = \lambda(\omega + u\omega')$ where λ is a proportionality factor. Substituting $-r\omega$ for ω'' in the left member of this relation yields $\omega' - ru\omega = \lambda(\omega + u\omega')$. This relation implies that $r^2u + 1 = 0$, since the points ω, ω' are linearly independent. The homogeneous coordinates of the two fixed points of the mapping are, therefore, given by the forms

$$(3.12) \qquad \zeta = \omega + i\omega'/\sqrt{r}, \qquad \eta = \omega - i\omega'/\sqrt{r}.$$

Taylor's development for $\omega(t + h)$ along an $r =$ const. curve yields, in view of (2.2)

$$(3.13) \qquad \begin{aligned} \omega(t + h) &= \omega(t)(1 - rh^2/2 + r^2h^4/4!\cdots) \\ &\quad + \omega'(t)(h - rh^3/3! + r^2h^5/5!\cdots) \\ &= \omega\cos\sqrt{r}h + \omega'(\sin\sqrt{r}h)/r. \end{aligned}$$

Since the above series represent integral functions, the representation is valid for all real values of h. Solving (3.12) for ω and ω' in terms of ζ, η and substituting the expression for them in (3.13) yields

$$(3.14) \qquad \omega(t + h) = \zeta\exp(-i\sqrt{r}h) + \eta\exp(i\sqrt{r}h).$$

Since the general homogeneous coordinates of the points ζ, η are independent of h, these points form a convenient basis for the definition of local homogeneous coordinates. The local homogeneous coordinates of the point $\omega(t + h)$ with respect to this basis will be the coefficients of ζ and η in the form (3.14), and the local non-homogeneous coordinates of this point will be $u = \exp(2i\sqrt{rh})$. *The double points* ζ, η *of a homography which maps a curve* C *upon itself lie on the curve if the mapping is loxodromic or hyperbolic; they coincide in a point of* C *if the mapping is parabolic; or they do not lie on the curve if the mapping is elliptic.*

It has been shown that a mapping of a curve C upon itself by a transformation (3.2) is represented by a parametric displacement $\bar{t} = t + h$. If the displacement is defined in another parametric system by $\bar{v} = v + k$, it follows that v and t are related by an equation of the form

$$v = ct + d \tag{3.15}$$

where c and d are constants. For, let v be a continuous function of t with a derivative defined for each value of t; then $v(t + h) - v(t) = k(h)$, where $\bar{v}(t) = v(t + h)$. Dividing the members of this equation by h and letting h tend to zero yields the relation $v'(t) = \lim_{h \to 0} k(h)/h$. Since the right member is independent of t, $v'(t) = \text{constant}$, and therefore v is given by (3.15).

It may also be shown, although the proof will be omitted here, that if two distinct displacements of the points of a curve, represented in different parameters by $\bar{t} = t + h$ and $\bar{v} = v + h$, result from two homographic mappings of the z plane which have the same pair of fixed points ζ, η the parameters v, t are related by an equation of the form (3.15).

4. A homographic metric. To establish a suitable local reference frame for the definition of homogeneous coordinates of a moving point on a curve C, let $\omega(t)$ denote normal coordinates of the point. Consider the $r = \text{constant}$ curve which is tangent to C at ω. A mapping of the plane which maps this curve upon itself has two fixed points given by

$$\zeta = \omega + i\omega'/\sqrt{r}, \qquad \eta = \omega - i\omega'/\sqrt{r}.$$

The three points ω, ζ, η determine a circle Γ, except when ζ and η coincide. The harmonic conjugate of ω with respect to ζ and η is the point ω'. The *points* ω, ω' *form a suitable basis for a local coordinate system since homogeneous coordinates of any point in the plane are expressible in the form* $\omega + u\omega'$ *where* u *is a complex number.* If ζ and η coincide, they coincide in the point ω'. It may be easily shown that *the circle* Γ *determined by* ω, ζ, η *makes with* C *at* ω *an angle equal to the angle of the complex number* $i/\sqrt{r}$ (the proof will be omitted). This angle is $\pi/2$ if r is real and positive;

it is zero if r is real and negative. The circle Γ is, therefore, tangent to C at ω if at ω the function $r(t)$ is real and positive.

The homographic element of arc length of a curve C from a point $\omega(t)$ to a neighboring point ω_1, given by $\omega_1 = \omega(t + dt)$ will be defined by the invariant differential form

$$d\sigma = 2\sqrt{|\,r\,|}\,dt. \tag{4.1}$$

Geometric interpretations will now be given for the forms $d\sigma$ and $d\sigma^2$. The element $d\sigma$ is given by the relations

$$d\sigma = 2\,|\,P(\omega\omega'\omega_1\zeta)\,| = |\,P(\log(\zeta\eta\omega\omega_1))\,|\,, \tag{4.2}$$

in which the principal value of the logarithm is used and P is used to denote the principal parts of the cross-ratio $(\omega\omega'\omega_1\zeta)$ and the logarithm of the cross-ratio $(\zeta\eta\omega\omega_1)$. These results may be verified by making use of the local non-homogeneous coordinates 0, ∞, dt, $i\sqrt{r}$, and $-i\sqrt{r}$ of the points ω, ω', ω_1, ζ and η, respectively. Let $\bar{\Gamma}$ denote the circle determined by the points ω_1, ζ, η and if this circle is distinct from the circle Γ it intersects Γ in a point which will be denoted by ω_2. The general homogeneous coordinates of ω_2 are found to be given by the relation $w_2 = \omega - \omega'/|\,r\,|\,dt$. If, however, the circles $\bar{\Gamma}$ and Γ, defined at ω, coincide, then $r < 0$ (real) and the expression for ω_2 defines the point $\omega + \omega'/rdt$ which is the harmonic conjugate of the point $\omega'_1 = \omega'(t + dt)$ with respect to the points ω, ω'. In either case the point ω_2 may be used to complete geometric interpretations of $d\sigma^2$. These interpretations are given by the equations

$$4P(\log(\omega\omega_1\omega'\omega_2)) = -4P(\omega\omega'\omega_1\omega_2) = d\sigma^2. \tag{4.3}$$

Let the independent parameter t represent time. The motion of the point $\omega(t)$ along an $r =$ constant curve will be called loxodromic, hyperbolic, elliptic, or parabolic, according as $r(t)$ is a complex imaginary, a negative real, a positive real, or zero, respectively. A normal non-homogeneous coordinate u, called the *abscissa*, of a generic point $\omega(t)$ of an $r =$ const. curve will be determined for each type of motion. An angle θ will be determined in association with each type of motion in terms of which the abscissa of ω and, consequently, the homographic arc length of the curve will be simply expressed.

Case 1. $r = (\mu + i\nu)^2$, $\mu, \nu =$ const., $\nu \neq 0$. Normal coordinates of ω (independent solutions of (1.2)) are

$$\omega^1 = \exp(-i\sqrt{r}t), \qquad \omega^2 = \exp(i\sqrt{r}t),$$

where $\sqrt{r} = \mu + i\nu$. The abscissa of ω is defined by

$$u = \omega^2/\omega^1 = \exp\{2(-\nu + i\mu)t\}. \tag{4.4}$$

Let $\mu t = \theta$, so that

$$\log u = 2a\theta + 2i\theta \tag{4.5}$$

where $a = -\nu/\mu$. Hence

$$d\sigma = \sqrt{\mu^2 + \nu^2}\, dt = 2 \sec \phi d\theta \tag{4.6}$$

where $\phi = \operatorname{arc\,tan} \nu/\mu$. Homographic arc-length of C between points where $\theta = \theta_0$ and $\theta = \theta_1$ is therefore given by

$$\sigma = 2(\theta_1 - \theta_0) \sec \phi. \tag{4.7}$$

Case 2. $r = -k^2$. Normal coordinates of ω are $\omega^1 = \exp(-kt)$, $\omega^2 = \exp(kt)$. The abscissa of ω is defined by $u = \exp(2kt)$. Let $u = \tan \theta$, so that $2kt = \log \tan \theta$ and

$$d\sigma = d(\log \tan \theta) = 2 \csc 2\theta d\theta. \tag{4.8}$$

Homographic arc-length of C between points where $\theta = \theta_0$ and $\theta = \theta_1$ is therefore given by

$$\sigma = \log(\tan \theta_1/\tan \theta_0). \tag{4.9}$$

In case the curve C is a "non-euclidean" straight line in the half-plane of Poincaré, the formula (4.7) defines the Poincaré metric of C (see [1, pp. 37-39] and [8, pp. 7-8].)

Case 3. $r = k^2$. Normal coordinates of ω are $\omega^1 = \cos kt$, $\omega^2 = \sin kt$. The abscissa of ω is defined by $u = \tan kt$. Let $u = \tan \theta$, so that $kt = \theta$ and

$$d\sigma = 2d\theta. \tag{4.10}$$

The homographic arc-length of C between the points where $\theta = \theta_0$ and $\theta = \theta_1$ is therefore

$$\sigma = 2(\theta_1 - \theta_0). \tag{4.11}$$

Case 4. $r = 0$. Normal coordinates of ω are $\omega^1 = 1$, $\omega^2 = t$. The local non-homogeneous coordinate of ω is defined by $u = t$. Here $d\sigma = 0$, but a geometric interpretation may be given for a parameter θ such that

$$u = \tan \theta. \tag{4.12}$$

The geometric meanings of the parameters θ satisfying the above relations in the various cases will now be given. Consider a pencil of circles passing through the fixed points ζ, η of a loxodromic mapping corresponding to case 1. Let any selected one of these circles C_1 serve as a basis; the angle at ζ between the circle C_1 and the circle of the pencil which passes through ω is a parameter θ which satisfies equation (4.5). To prove this assertion, note that the equation (4.5) is the polar equation of a logarithmic spiral one of whose points is infinite and the other of which is the pole of the coordinate system, the angle θ being the polar angle of the point ω. The interpretation of an angle θ for the case of an $r =$ const. curve (Case 1.) with finite fixed points ζ, η is obtained immediately on subjecting the logarithmic spiral and the $\theta =$ const. lines of the polar coordinate system to a suitable homography which transforms the pole and infinite point into the points ζ, η and transforms the initial line into the circle C_1.

To obtain a geometric interpretation of a parameter θ in terms of which $d\sigma$ is given by (4.8), construct the circle C_1 through the points ζ, η orthogonal to C. Select on C_1 an arbitrary point I distinct from ζ and η. Construct the circle C_ω determined by the points I, ζ, ω. *The angle between* C_1 *and* C_ω *at* ζ *is a parameter* θ *in terms of which* $d\sigma$ *is given by* (4.8). To prove this let θ denote this parameter and calculate the non-euclidean element $d\sigma$ in terms of θ. Transform the plane by a homography which transforms ζ, η into given points $\bar{\zeta}$, $\bar{\eta}$ and the point I into the point $\bar{I}$ at infinity. The transform $\bar{C}$ of C is, therefore, a circle orthogonal to the straight line transform $\bar{C}_1$ of C_1, and the transform $\bar{C}_\omega$ of C_ω is a straight line (passing through the points $\bar{I}$, $\bar{\zeta}$, $\bar{\omega}$) which makes with $\bar{C}_1$ the angle θ. Let Γ_1 denote the circle determined by I, ζ, ω_1. The transform $\bar{\Gamma}_1$ of this circle is a straight line through $\bar{\zeta}$ making with $\bar{C}_1$ the angle $\theta + d\theta$. Since the element $d\sigma$ is an invariant of homographies, from equations (4.2) it follows that

$$d\sigma = P(\log(\bar{\zeta}\bar{\eta}\bar{\omega}\bar{\omega}_1)) = \log(\bar{\xi}\bar{\omega})(\bar{\eta}\bar{\omega}_1)/(\bar{\eta}\bar{\omega})(\bar{\zeta}\bar{\omega}_1) = d(\log\tan\theta).$$

This completes the proof.

A parameter θ in terms of which $d\sigma$, for Case 3, is defined by (4.10) may be geometrically characterized as follows A circle C at whose points $r = k^2$ is orthogonal to all of the circles of a pencil which pass through the fixed points ζ and η. Let C_1 be any selected circle of the pencil and let Γ denote the circle of the pencil which passes through the point ω. The angle which the circle Γ makes with the circle C is equal to 2θ, where θ is a parameter of the type sought. For the proof transform by a homography the point η into the point $\bar{I}$ at infinity; the pencil of circles transforms into a pencil of lines with center at $\bar{\zeta}$ orthogonal to the transform $\bar{C}$ of C. It follows that the homographic element $d\sigma$ is given by

$$d\sigma = \mid P(\log(\bar{\zeta}\bar{I}\bar{\omega}\bar{\omega}_1) \mid = \mid P(\log(\bar{\zeta}\bar{\omega})/\,(\bar{\zeta}\bar{\omega}_1))\mid = 2d\theta.$$

A parameter θ in terms of which the abscissa of ω is given by the relation (4.12) may be characterized as follows: Select on the circle C an arbitrary point I and construct a circle C_1 which passes through ζ and I and is orthogonal to C. Select on C_1 any other point R and through the point ω of C draw the circle Γ determined by the points R, I, and ω. The angle formed at R by the two circles C_1 and Γ is a parameter θ in terms of which u is given by the relation (4.12). The proof, which is similar to those of the former cases, will be omitted.

A generic point ω of an $r = \text{constant}$ curve is given by

$$\omega = \zeta + \eta \exp(2i\sqrt{r}t) \tag{4.13}$$

in which ζ, η denote the double points of a mapping of the curve upon itself. The proofs of the following well known properties, by the methods of the present paper, may be readily supplied by the reader on making use of the fact that a homographic mapping is conformal. *A homography other than the identity which maps a curve upon itself maps the curves of a one parameter family upon themselves. If the mapping is* (1) *loxodromic,* (2) *hyperbolic,* (3) *elliptic, or* (4) *parabolic, the family consists of* (1) *non-degenerate loxodromes,* (2) *circles which pass through the double points of the mapping,* (3) *the orthogonal trajectories of the circles which pass through the double points of the mapping, or* (4) *the circles tangent to a given circle at the double point of a parabolic mapping, respectively.*

5. The geometry of the absolute.

In keeping with the geometric characterization given by (4.2) for the element of homographic arc length of a curve, let the homographic distance between two points P, Q with respect to an absolute (ζ, η) be defined by

$$\overline{PQ} = \mid \log(\zeta\eta PQ) \mid \tag{5.1}$$

in which the principal value of the logarithm is used. In the geometry of the absolute (ζ, η) in which the metric is defined by (5.1) *a "straight" line will be defined to be a curve whose homographic arc length between any two of its points P, Q is equal to the homographic distance between these two points.* To avoid confusing these lines with ordinary straight lines they will be called *s-lines.* To find the s-lines of the plane let a generic point ω of an s-line be given by

$$(5.2) \qquad \omega = \zeta + e^{w}\eta \qquad\qquad w = \tau + i\theta$$

where τ is a function of θ to be determined. Let $\tau(\theta_0)$ be denoted by τ_0; the homographic distance between two points ω_0, ω now assumes the form

$$(5.3) \qquad |\log(\zeta\eta\omega_0\omega)| = [(\tau - \tau_0)^2 + (\theta - \theta_0)^2]^{\frac{1}{2}}.$$

Hence, the condition that the locus of ω be an s-line is that

$$(5.4) \qquad [(\tau - \tau_0)^2 + (\theta - \theta_0)^2]^{\frac{1}{2}} = \int_{\omega_0}^{\omega} (d\tau^2 + d\theta)^{\frac{1}{2}}$$

for any two points ω_0, ω of the locus. If $\theta = \text{const.}$, differentiating equation (5.4) with respect to θ, squaring both sides of the resulting equation, and clearing of fractions yields the equation $(\theta - \theta_0)d\tau = (\tau - \tau_0)d\theta$ whose general solution is of the form

$$(5.5) \qquad \tau = a\theta + b, \qquad\qquad a, b = \text{const.}$$

Let ρ denote $|e^{w}|$, then ρ is given by

$$(5.6) \qquad \rho = ce^{a\theta}, \qquad\qquad a, c = \text{const.}$$

Since equation (5.2) with ρ given by (5.6) defines a two parameter family of curves, it is clear that only one s-line passes through any two given points which are distinct from the points of the absolute, and only one s-line passes through a given point (distinct from ζ or η) in a given direction. The s-lines thus determined by (5.2) and (5.6) are in general loxodromes, and along an s-line where $\theta \neq$ a constant these loxodromes are non-degenerate. The s-line along which $a = 0$ are the orthogonal trajectories of the circles $\theta_0 =$ constant which pass through the points ζ and η.

Along an arc of a $\theta = \text{const.}$ circle the condition (5.4) assumes the form

$$\tau - \tau_0 = \int_{\tau_0}^{\tau} d\tau$$

which is satisfied identically in τ. Since the values of τ which correspond to the points ζ and η of the absolute are $-\infty$ and $+\infty$ respectively, an arc of a circle contained between ζ and η is an entire s-line. The following general result may now be stated: *An s-line is a curve which makes a constant angle with the circles which pass through the points of the absolute. If the* *angle is not a right angle the s-line is a loxodrome whose asymptotic points are ζ and η or an arc of a circle whose endpoints are ζ and η.* An $r =$ const. curve is an s-line with respect to the absolute formed by the fixed points ζ, η of a mapping of the curve upon itself.

In the geometry of the metric (5.1) the points ζ, η of the absolute are the points "at infinity," that is, the homographic distance from either ζ or η to any other point in the plane is infinite, whereas the homographic distance between any two points distinct from ζ and η is finite. The numbers θ, τ will be called normal curvilinear coordinates of the point ω with respect to the absolute ζ, η. It is clear that a set of coordinates θ, τ determines by (5.2) one and only one point ω. The transformation (5.2) maps the lines $\tau = -\infty$ and $\tau = +\infty$ into the points ζ and η, respectively, but otherwise maps the entire finite w-plane conformally upon the points of the z-plane. A straight line of the w-plane corresponds to an s-line of the z-plane, and (5.3) shows that the homographic distance between any two points ω_0, ω of the z-plane is equal to the ordinary distance between the corresponding points in the w-plane. The geometry of the z-plane (except at the points ζ and η ("at infinity") is therefore *euclidean.* This, of course, implies that the sum of the interior angles of any triangle formed by s-lines is equal to π radians (a result which can also be easily obtained by making use of the fact that an s-line cuts at a constant angle the circles which pass through the points ζ and η). The curious case of an s-line triangle with one vertex ω_1 superposed upon another ω_0 occurs if θ_1 differs from θ_0 by an integral multiple of 2π radians; in this case the side $\omega_0\omega_1$ is a complete circumference of a circle.

Since the arcs $\zeta\eta$ of circles and the family of orthogonal trajectories are the families of $\theta =$ const. curves and $\tau =$ const. curves, respectively, these families are families of *equidistants,* in the geometry of the absolute, in which distances between members of one family are measured along the members of the other family. The members of each of these families will be called *parallels* with respect to the absolute. Moreover, in general, distinct curves which make equal constant angles with all of the $\theta =$ constant curves will be called parallel; since such curves cannot intersect in any point (ρ, θ) distinct from the points of the absolute.

A congruent transformation in the geometry of this metric is a transformation which preserves the points of the absolute; two figures which correspond in a congruent transformation are called congruent figures. The mapping of figures in the z-plane which are congruent with respect to homographies by the transformation (5.2) produces figures in the w-plane which are congruent with respect to the ordinary rigid displacements. It is, therefore, logical to call the homographies which preserve the points of the absolute (the points ζ, η "at infinity") displacements with respect to the homographic metric.

UNIVERSITY OF KANSAS.

BIBLIOGRAPHY.

[1] E. Cartan, *Leçons sur la théorie des espaces à connexion projective*, Paris, Gauthier-Villars (1937).

[2] W. F. Osgood, *Lehrbuch der Funktionentheorie*, Leipzig, B. G. Teubner (1912).

[3] H. Liebmann, "Beitrage zur Inversionsgeometrie der Kurven," *Sitzungsberichte der Bayerischen Akademie der Wissenschaften zur München*, Heft 1 (1923).

[4] T. Kubota, "Beitrage zur Inversionsgeometrie," *Tohoku Imperial University, Science Reports*, vol. 13 (1924-25), pp. 243-251.

[5] Jusaku Maeda, "Geometrical meanings of the inversion curvature of a plane curve," *Japanese Journal of Mathematics*, vol. 16 (1940), pp. 177-232.

[6] Frank Morley, "On differential inversive geometry," *American Journal of Mathematics*, vol. 48 (1926), pp. 144-146.

[7] B. C. Patterson, "The differential invariants of inversive geometry," *American Journal of Mathematics*, vol. 50 (1928), pp. 553-568.

[8] H. Poincaré, "Théorie des groupes fuchsiens," *Acta Mathematica*, vol. 1 (1882), pp. 1-62.

[9] P. O. Bell, "A characterization of the group of homographic transformations," *Bulletin of the American Mathematical Society*, vol. 47 (1941), pp. 488-493.

DOUBLE VECTOR SPACES OVER DIVISION RINGS.*

By G. Hochschild.

Introduction. Let M be a (left) vector space over the division ring D. Let E be an arbitrary ring with an identity element, and suppose there is given an antihomomorphism of E into the ring of D-linear transformations in M, such that the identity element of E is mapped into the identity transformation. If we denote by E^* the ring of right multiplications $x \to e^*\{x\} = xe$ in E then the mapping $e \to e^*$ is an anti-isomorphism of E onto E^*, and the situation described above amounts to having a representation of E^* as D-linear transformations in M.

Under these circumstances, it is convenient to think of M as having simultaneously the structure of a left D-space and that of a right E-space, and to indicate the corresponding scalar operations as $m \to d \cdot m$, for $d \varepsilon D$, and as $m \to m \cdot e$, for $e \varepsilon E$. Our above requirements may then all be absorbed in the statement that the " $\cdot$ products" should behave like ordinary products. We shall say then that M has the structure of a (D, E)-space.

The theory of (D, E)-spaces may, of course, be subsumed under the theory of matrices with coefficients in D. A study of matrices with coefficients in a division ring which is relevant here was made by R. Brauer in 1941,[1] but this overlaps only slightly with what we propose to do.

In the special case where D and E both coincide with a commutative field F, the study of (F, F)-spaces amounts to the same as Jacobson's theory of self-representations.[2] Jacobson has shown that the self-representations provide a generalization of Galois theory to general field extensions, not necessarily normal, or even separable. Part of our program is to consider this aspect in the non-commutative case.

The Galois theory for division rings has recently been developed by N. Jacobson [3] and, independently, by H. Cartan.[4] Their chief tool is a theorem—

---

* Received January 30, 1948; revised July 20, 1948.

[1] R. Brauer, "On sets of matrices with coefficients in a division ring," *Transactions of the American Mathematical Society*, vol. 49 (1941), pp. 502-548.

[2] N. Jacobson, "An extension of Galois theory to non-normal and non-separable fields," *American Journal of Mathematics*, vol. 66 (1944), pp. 1-29.

[3] N. Jacobson, "A note on division rings," *ibid.*, vol. 69 (1947), pp. 27-36.

[4] H. Cartan, "Théorie de Galois pour les corps non commutatifs," *Annales Scientifiques de l'Ecole Normale Supérieure*, vol. 65 (1948), pp. 60-77.

due to Jacobson and to N. Bourbaki—on rings of additive endomorphisms of a division ring. We shall state this theorem in a slightly generalized form in **2** which also contains a completely elementary proof. In **3**, we give a brief indication of how this theorem is used in order to establish the main results of the Galois theory.

An important feature of this approach is that the Galois correspondence between division subrings of a division ring D and groups of automorphisms of D is derived from the correspondence between division subrings K of D and the rings of all K-linear transformations in D. This last correspondence may be regarded as a generalization of the usual Galois correspondence. The rôle played by a single automorphism is then taken over by an indecomposable space of additive endomorphisms of D.

As Jacobson has shown, the full operator ring, or the "composite," of a self-representation of a field F is determined to within an equivalence by a certain space of additive endomorphisms of F, its so-called relations space, and it is precisely this fact which ties up the theory of self-representations with the Galois theory. In the non-commutative case, there arises the difficulty that the composite of a representation may easily fail to have the necessary finiteness properties, and cannot be made a part of an adequate theory. There is, however, another aspect of the composite which is capable of a satisfactory generalization.

Abstractly, a composite of two rings D and E with identity elements is a ring R which is generated by homomorphic images D' and E' of D and E, respectively, in which every element of D' commutes with every element of E', and where the identity elements of D and E are both mapped into the identity element of R. Clearly, R carries the structure of a (D, E)-space in the natural way, and R is generated as a (D, E)-space from its identity element. This has led us to introduce in **5** the notion of a relative cyclic (D, E)-space, i. e., a (D, E)-space M in which we have singled out an element m such that $D. m. E = M$. To such a space we attach a certain space of additive homomorphisms of E into D, its relations space, which coincides with Jacobson's relations space in the case where M is the composite of a field F with itself.

In **5** and **6**, we establish the main facts concerning the correspondence between relative cyclic spaces and their relations spaces. In **7**, we apply these facts to the particularly interesting case where E is a division ring and D a division subring of E. In this case, the product of two relative cyclic (D, E)-spaces is always defined and leads to a notion of closure for such spaces. A space is closed if and only if its relations space is a ring. Furthermore, the closed spaces are exactly the Kronecker product spaces $D \times E$, taken with respect to some division subring of D.

There is another type of double vector space which is useful in studying extensions of a division ring K. This is the K-regular space defined in **4**. The purpose of the restriction to K-regularity is to make the transformations corresponding to the elements of K as simple as possible without losing too much generality. For the case where D is a finite normal extension of K, and E is a division ring between K and D, we obtain in **4** a complete theory of K-regular (D, E)-spaces which reflects the main facts of the Galois theory for D over K. In this case, the situation is so simple that we have no need of relations spaces.

I wish to express my thanks to Prof. R. Brauer whose critical comments on an earlier version of this paper have been very valuable to me in the revision and to whom I am indebted for a number of improvements, notably in **2** and **4**.

1. Kronecker products. Let D be a division ring. By a vector space over D we shall mean an additive group A, together with an isomorphism ϕ of D into the ring of endomorphisms of A, such that—if 1 is the identity element of D—$\phi\{1\}$ is the identity mapping in A. If $d \varepsilon D$ and $a \varepsilon A$ we shall write $d \cdot a$ for $(\phi\{d\})\{a\}$. The $\cdot$ operation has then the formal properties of a multiplication.

What we have just described is frequently called a left vector space over D. The notion of a right vector space is obtained by replacing the above isomorphism ϕ by an anti-isomorphism. In order to avoid confusion, we shall use the phrase "vector space" only to denote a left vector space. Similarly, by the dimension of a space, we shall always mean its dimension as a left vector space. We shall absorb the notion of a right vector space in our terminology by means of the following device: Let D^* denote the ring of right multiplications in D. We make use of the anti-isomorphism $d \rightarrow d^*$ of D onto D^* which we have defined at the beginning of the introduction in order to regard a right vector space B over D as a left vector space over D^*. We shall frequently write $b \cdot d$ for $d^* \cdot b$, i. e., we shall adhere to the usual formalism for right vector spaces.

From a vector space A over D and a vector space B over D^* we can construct a certain additive group, called the Kronecker product $B \times A$. We need not give such a construction here, since quite a variety of suitable ones are well known. We shall merely enumerate certain fundamental properties which characterize the Kronecker product.

(1) To every pair (b, a), where $b \varepsilon B$ and $a \varepsilon A$, there corresponds an element $b \times a$ of $B \times A$.

(2) The elements $b \times a$ generate $B \times A$, i. e., every element of $B \times A$ is the sum of a finite number of such elements.

(3) $(b_1 \pm b_2) \times a = b_1 \times a \pm b_2 \times a$.

(4) $b \times (a_1 \pm a_2) = b \times a_1 \pm b \times a_2$.

(5) $b \cdot d \times a = b \times d \cdot a$, for $d \varepsilon D$.

(6) If (a_γ) ($\gamma \varepsilon$ some finite set Γ) is a set of D-linearly independent elements of A, and if $b(\gamma)$ are elements of B, not all equal to 0, then $\sum_{\gamma \varepsilon \Gamma} b(\gamma) \times a_\gamma \neq 0$.

(7) If (b_δ) is a finite set of D^*-linearly independent elements of B, and if $a(\delta)$ are elements of A, not all 0, then $\sum b_\delta \times a(\delta) \neq 0$.

Let S be a D-linear transformation in A. Then S induces in $B \times A$ an endomorphism S' such that $S'\{b \times a\} = b \times S\{a\}$. Similarly, if T is a D^* linear transformation in B then T induces in $B \times A$ an endomorphism T' such that $T'\{b \times a\} = T\{b\} \times a$. We shall apply these facts to the following situation: Let B be an (E, D)-space and A a (D, F)-space, where E and D are division rings. Then we may construct the Kronecker product $B \times A$ with respect to D as above, and we see that now $B \times A$ carries the structure of an (E, F)-space such that, for $e \varepsilon E$ and $f \varepsilon F$, we have $e \cdot (b \times a) = e \cdot b \times a$, and $(b \times a) \cdot f = b \times a \cdot f$.

2. Additive homomorphisms of division rings. Let S be an arbitrary set, D a division ring. The additive group of all mappings of S into D may be regarded as a vector space over D^*, the ring of right multiplications in D, where the scalar multiple of a mapping T by an element $d^* \varepsilon D^*$ is defined as the ordinary product d^*T (i. e., T followed by d^*). We shall require the following elementary lemma:

LEMMA 2.1. *Let $T_1, \cdots, T_n$ be a set of mappings of the set S into the division ring D which are linearly independent with respect to D^*. Then there exist n linearly independent D^*-linear combinations $U_1, \cdots, U_n$ of the T_i and elements $s_1, \cdots, s_n$ in S, such that $U_i\{s_j\} = \delta_{ij}$, where $\delta_{ii} = 1$ and $\delta_{ij} = 0$ if $i \neq j$.*

Proof. Suppose we have already found $s_1, \cdots, s_k$ in S and independent D^*-linear combinations $T_1^{(k)}, \cdots, T_n^{(k)}$ of the T_i, such that $T_i^{(k)}\{s_j\} = \delta_{ij}$, for $1 \leqq i \leqq n$ and $1 \leqq j \leqq k$. Here, we wish to include also the case $k = 0$, where we have no elements in S, and where $T_i^{(0)}$ is to stand for T_i. Then, if $k < n$, there is an element $s_{k+1} \varepsilon S$ such that $T_{k+1}^{(k)}\{s_{k+1}\} \neq 0$. We set $T_{k+1}^{(k+1)} = (T_{k+1}^{(k)}\{s_{k+1}\}^{-1})^* T_{k+1}^{(k)}$, so that $T_{k+1}^{(k+1)}\{s_{k+1}\} = 1$. For every $j \neq k+1$, we set $T_j^{(k+1)} = T_j^{(k)} - (T_j^{(k)}\{s_{k+1}\})^* T_{k+1}^{(k+1)}$. Then we have $T_i^{(k+1)}\{s_j\} = \delta_{ij}$, for $1 \leqq i \leqq n$ and $1 \leqq j \leqq k+1$, and the $T_i^{(k+1)}$ are still linearly independent. Proceeding in this fashion, starting from the case $k = 0$, we finally obtain $s_1, \cdots, s_n$ in S and $U_i = T_i^{(n)}$ which satisfy the requirements of the lemma.

Now let D be a division subring of the division ring E. The additive homomorphisms of E into D constitute a ring which is also a (D^*, E^*)-space, all operations being defined in the natural fashion as ordinary compositions of mappings. The most important subrings of this ring are those which are also (D^*, E^*)-subspaces:

Definition 2.1. Let A be a subring of the ring of all additive homomorphisms of the division ring E into the division subring D of E. A is called complete if $A \neq (0)$, and if $D^*AE^* = A$.

If A is complete then the only element of E which is mapped into 0 by all the elements of A is 0. In fact, if we had $0 \neq e_0 \varepsilon E$ such that $U\{e_0\} = 0$ for all $U \varepsilon A$ then, for every $V \varepsilon A$ and every $e \varepsilon E$, we should have $V\{e\} = (V(e_0^{-1}e)^*)\{e_0\} = 0$, contradicting our hypothesis that $A \neq (0)$.

The fundamental result concerning complete rings is the following:

THEOREM 2.1 (Jacobson-Bourbaki). *Let A be a complete subring of the ring of additive homomorphisms of E into $D \subseteq E$. Denote by K the division subring of D which consists of all elements $d \varepsilon D$ for which $U\{de\} = dU\{e\}$, for all $e \varepsilon E$ and all $U \varepsilon A$* (i. e., whose left multiplications $e \rightarrow de$ commute with all the elements of A). *Then if either E is of finite dimension $[E:K]$ over K, or if A is of finite dimension $[A:D^*]$ over D^*, we have $[E:K] = [A:D^*]$, and A coincides with the ring of all K-linear mappings of E into D.*

Proof. If there are n D^*-linearly independent elements of A we can apply lemma 2.1 to find elements $e_1, \cdots, e_n$ in E and elements $U_1, \cdots, U_n$ in A such that $U_i\{e_j\} = \delta_{ij}$. Then the e_j are evidently linearly independent over K. Hence if E is finite over K then A must be finite over D^*. Thus, we may suppose that $[A:D^*] = n$, and that the above U_i constitute a basis for A over D^*.

Then, if U is an arbitrary element of A, we must clearly have $U = \sum_{i=1}^{n} U\{e_i\}^* U_i$.

Let $V \varepsilon A$, $e \varepsilon E$, and set $U = Ve^* U_j$. Then $U \varepsilon A$, and the last relation gives

$$Ve^* U_j = \sum_{i=1}^{n} (V\{U_j\{e_i\}e\})^* U_i = V\{e\}^* U_j.$$

Hence, for every $e_0 \varepsilon E$, $V\{U_j\{e_0\}e\} = U_j\{e_0\}V\{e\}$, which means that $U_j\{E\} \subseteq K$.

Now, for $e \varepsilon E$, set $e' = e - \sum_{k=1}^{n} U_k\{e\}e_k$. Then $U_i\{e'\} = 0$, for all i, whence $U\{e'\} = 0$, for all $U \varepsilon A$. Hence $e' = 0$, i. e., $e = \sum_{k=1}^{n} U_k\{e\}e_k$. Thus, $e_1, \cdots, e_n$ is a basis for E over K, so that $[E:K] = n = [A:D^*]$. Finally, if T is any K-linear mapping of E into D we have clearly

$$T\{e_i\} = (\sum_{k=1}^{n} (T\{e_k\})^* U_k)\{e_i\},$$

whence

$$T = \sum_{k=1}^{n} (T\{e_k\})^* U_k \,\varepsilon\, A.$$

Since every element of A is evidently a K-linear mapping, our proof is complete.

3. Galois theory.

Just as in the case of fields, the following property of "normality" of an extension is fundamental in the Galois theory:

Definition 3.1. Let D be a division ring containing the division ring K. We say that D is normal over K if every element of D which is left fixed by every automorphism of D over K belongs to K.

The results of the Galois theory are obtained by applying Theorem 2.1 to the rings of endomorphisms of the additive group of D which are generated by the right multiplications and the elements of a group of automorphisms of D. An important fact which we shall make use of later is the following:

Let D be normal and of finite dimension n over K. Then the ring of all K-linear transformations in D has a basis over D^* which consists of n automorphisms of D over K.

The ring of all additive endomorphisms of D has the structure of a (D^*, D^*)-space in a natural manner. If g is an automorphism of D then $gd^* = g\{d\}^* g$, whence we see that $D^* g$ is a (D^*, D^*)-subspace of the space of all additive endomorphisms.

Conversely, if M is any (D^*, D^*)-space of additive endomorphisms of D which is of dimension 1 over D^* then M contains an element g which is a ring automorphism of D, and $D^*g = M$. In fact, g is the unique element of M which maps 1 into 1.

If D is finite and normal over K it can be shown that every indecomposable (D^*, D^*)-space of K-linear transformations in D is of dimension 1 over D^* and hence is generated by a unique automorphism of D over K.[5]

Noting further that $(D^*g_1)(D^*g_2) = D^*g_1g_2$, we see that if we regard the (D^*, D^*)-spaces of additive endomorphisms of D as generalizations of sets of automorphisms of D the ordinary multiplication, in the ring of all additive endomorphisms, of two such subspaces is the analogue of the multiplication of sets of automorphisms. An indecomposable (D^*, D^*)-space of additive endomorphisms is the proper analogue of a single automorphism. It must be kept in mind, however, that, while the product of two automorphisms is again a single automorphism, the product of two indecomposable spaces of additive endomorphisms need, in general, not be indecomposable.

4. Double vector spaces over normal division rings.

Definition 4. 1. Let D and E be division rings, each containing the division ring K. A (D, E)-space V is said to be K-regular if there exists a subset V_0 of V such that $D \cdot V_0 \cdot E = V$ and $k \cdot v = v \cdot k$, for all $k \in K$ and all $v \in V_0$.

Suppose that $D \supseteq E \supseteq K$, and let g be an isomorphism of E into D which leaves the elements of K fixed. Then we construct a K-regular (D, E)-space as follows:

The underlying additive group is a copy, D' say, of that of D. We denote by $d \rightarrow d'$ an isomorphism of the additive group of D onto D'. The scalar operations in D' are defined by setting $d_1 \cdot d' = (d_1 d)'$, and $d' \cdot e = (dg\{e\})'$.

Clearly, the set consisting of the element $1'$ alone satisfies the requirements for V_0 above. Thus we have a K-regular (D, E)-space which we shall denote $V(g)$.

If g_1 and g_2 are two isomorphisms of E into D which leave the elements of K fixed then $V(g_1)$ and $V(g_2)$ are (D, E)-operator isomorphic if and only if there is an inner automorphism a of D such that $g_1 = ag_2$.

In fact, denote the additive groups of $V(g_1)$ and $V(g_2)$ by D' and D'', respectively. If $g_1\{e\} = d_0 g_2\{e\} d_0^{-1}$, with some fixed d_0 for all $e \varepsilon E$, we define a mapping ϕ of D' onto D'' by setting $\phi\{d'\} = (dd_0)''$. Then

---

[5] This follows, for instance, from the results in §§ 4, 5 and 6 of this paper.

$$d_1 \cdot \phi\{d'\} = (d_1 d d_0)'' = \phi\{(d_1 d)'\} = \phi\{d_1 \cdot d'\},$$

and

$$\phi\{d'\} \cdot e = (d d_0 g_2\{e\})'' = (d g_1\{e\} d_0)'' = \phi\{(d g_1\{e\})'\} = \phi\{d' \cdot e\},$$

i. e., ϕ is an operator homomorphism. Moreover, it is evident that the kernel of ϕ is zero and that ϕ maps $V(g_1)$ onto $V(g_2)$.

Conversely, if ϕ is an operator isomorphism of $V(g_1)$ onto $V(g_2)$, and if $\phi\{1'\} = d_0''$, then we deduce that $g_1\{e\} = d_0 g_2\{e\} d_0^{-1}$. In fact, $\phi\{1'\} \cdot e = \phi\{1' \cdot e\} = \phi\{g_1\{e\}'\}$, i. e., $d_0'' \cdot e = g_1\{e\} \cdot \phi\{1'\} = (g_1\{e\} d_0)''$, or $(d_0 g_2\{e\})'' = (g_1\{e\} d_0)''$, whence the result.

Now let us suppose that D is finite and normal over K and consider the Kronecker product $D \times E$ with respect to K, where $D \supseteq E \supseteq K$.

This is a K-regular (D, E)-space. In fact, $D \times E = D \cdot (1 \times 1) \cdot E$, and $k \cdot (1 \times 1) = k \times 1 = 1 \times k = (1 \times 1) \cdot k$, for every $k \varepsilon K$.

Let $[D:K] = n$. As we have remarked in **3**, there exist n automorphisms $g_1, \cdots, g_n$ of D over K which form a basis over D^* for the space of all K-linear transformations in D. Since every K-linear mapping of E into D can evidently be extended to a K-linear transformation in D, the restrictions $\bar{g}_i$ of the g_i to E must contain a basis over D^* for the space of all K-linear mappings of E into D. We may suppose that $\bar{g}_1, \cdots, \bar{g}_m$ is such a basis. Then $m = [E:K]$.

For each index i, $1 \leqq i \leqq m$, there exists a homomorphism π_i of the additive group of $D \times E$ onto that of $V(\bar{g}_i)$—whose elements we shall now denote by $d^{(i)}$—which is characterized by the relations $\pi_i\{\sum_a d_a \times e_a\} = \sum_a (d_a g_i\{e_a\})^{(i)}$, for all $d_a \varepsilon D$ and $e_a \varepsilon E$. Moreover, π_i is evidently a (D, E)-operator homomorphism of the (D, E)-space $D \times E$ onto the (D, E)-space $V(\bar{g}_i)$.

Now let M be the direct sum of the $V(\bar{g}_i)$, for $1 \leqq i \leqq m$, and let π be the mapping of $D \times E$ into M which is defined by setting $\pi\{z\} = (\pi_1\{z\}, \cdots, \pi_m\{z\})$, for every $z \varepsilon D \times E$. Evidently, π is a (D, E)-operator homomorphism. We claim that π is actually an isomorphism.

Let z be in the kernel of π. Then $\pi_i\{z\} = 0$, for each i. If $e_1, \cdots, e_m$ is a basis for E over K we can determine elements $d_{ij} \varepsilon D$ such that $\sum_{i=1}^{m} g_i\{e_k\} d_{ij} = \delta_{kj}$. The element z may be written in the form $z = \sum_{k=1}^{m} d_k \times e_k$.

We have then $\sum_{k=1}^{m} d_k g_i\{e_k\} = 0$, for each i, whence $\sum_{k,i} d_k g_i\{e_k\} d_{ij} = 0$, for each j; i. e., $d_j = 0$, and thus $z = 0$, which proves our assertion.

Finally, we note that the dimension of M over D is equal to m, the dimension of $D \times E$ over D. Hence π maps $D \times E$ onto M.

We are now in a position to prove the following theorem:

THEOREM 4. 1. *Let D be finite and normal over K, and let E be a division ring such that $D \supseteq E \supseteq K$. Then every K-regular (D, E)-space is a sum of simple K-regular subspaces. Every simple K-regular (D, E)-space is (D, E)-operator isomorphic with a $V(\bar{g})$, where $\bar{g}$ is the restriction to E of an automorphism g of D over K.*

Proof. Let N be any K-regular (D, E)-space. N is the sum of subspaces of the form $D \cdot n \cdot E$, where $n \varepsilon N$, and $k \cdot n = n \cdot k$, for every $k \varepsilon K$. Hence it suffices to prove our theorem in the case where $N = D \cdot n \cdot E$.

In that case, we can evidently define a (D, E)-operator homomorphism ψ of $D \times E$ onto N which is such that $\psi\{\sum_a d_a \times e_a\} = \sum_a d_a \cdot n \cdot e_a$. Hence there is also an operator homomorphism $h = \psi\pi^{-1}$ of $M = (V(\bar{g}_1), \cdots, V(\bar{g}_m))$ onto N. Let W be the kernel of h. By a standard argument of the theory of semisimple groups with operators, there are indices $i_1, \cdots, i_s$ such that M is the direct sum of W and $(V(\bar{g}_{i_1}), \cdots, V(\bar{g}_{i_s}))$. Then h induces an operator isomorphism of $(V(\bar{g}_{i_1}), \cdots, V(\bar{g}_{i_s}))$ onto N, which proves the first part of our theorem, since the $V(\bar{g}_i)$ are evidently simple. The rest follows by noting that if N is simple it must be of the form $D \cdot n \cdot E$, and s must be equal to 1.

Notice that theorem 4. 1 implies that every K-regular (D, E)-space is a *direct* sum of simple K-regular (D, E)-subspaces, as a standard argument, using Zorn's lemma, will show.

COROLLARY 1. *Let N be a K-regular (D, E)-space, where D and E are as above. Then there exists a K-regular (D, D)-space N' which, when regarded as a (D, E)-space, coincides with N.*

Proof. By the above remark, it suffices to show this for a simple (D, E)-space. By the second part of Theorem 4. 1, we may assume that the space is $V(\bar{g})$, where $\bar{g}$ is the restriction to E of an automorphism g of D over K. But then the (D, D)-space $V(g)$ evidently satisfies the requirements of our corollary.

Another consequence of Theorem 4.1 is the usual result on extensions of isomorphisms.[6]

Corollary 2. *Let D be normal and finite over K, E a division ring such that $D \supseteq E \supseteq K$. Then every isomorphism of E into D which leaves the elements of K fixed can be extended to an automorphism of D over K.*

Proof. Let a be such an isomorphism of E into D. Then the (D, E)-space $V(a)$ is simple and hence is (D, E)-operator isomorphic with a $V(\bar{g})$, where $\bar{g}$ is the restriction to E of an automorphism g of D over K. As we have seen above, this implies that there is an element $d \varepsilon D$ such that $a\{e\} = dg\{e\}d^{-1}$, for every $e \varepsilon E$. The automorphism $x \rightarrow dg\{x\}d^{-1}$ is the required extension of a to an automorphism of D over K.

A similar result is the following:

Corollary 3. *Let $D \supseteq E \supseteq K$, as above. Let a and b be two automorphisms of D over K, and let τ be an additive homomorphism of E into D such that $\tau\{K\} = (0)$, and $\tau\{e_1e_2\} = a\{e_1\}\tau\{e_2\} + \tau\{e_1\}b\{e_2\}$, for all $e_1, e_2 \varepsilon E$. Then there exists an element $d \varepsilon D$ such that $\tau\{e\} = a\{e\}d - db\{e\}$, for all $e \varepsilon E$.*

Proof. We define a K-regular (D, E)-space V of dimension 2 over D as follows: The additive group of V is the direct sum of two copies D' and D'' of that of D. For $d_1' \varepsilon D'$, $d_2'' \varepsilon D''$, and $d \varepsilon D$, we define $d \cdot (d_1', d_2'') = ((dd_1)', (dd_2)'')$, and, for $e \varepsilon E$, $(d_1', d_2'') \cdot e = ((d_1a\{e\})', (d_1\tau\{e\} + d_2b\{e\})'')$. It is easily verified that this defines in V the structure of a (D, E)-space. If V_0 is the subset consisting of the elements $1' = (1', 0)$ and $1'' = (0, 1'')$ we see that the requirements of definition 4, 1 are satisfied, so that V is K-regular.

Since D'' is a K-regular (D, E)-subspace of V, it follows from Theorem 4.1 that there exists a K-regular (D, E)-subspace W such that V is the direct sum of W and D''. Let us write, in accordance with this decomposition, $1' = w + d_1''$, where $w \varepsilon W$. Operating with $e \varepsilon E$, we obtain $a\{e\}' + \tau\{e\}'' = w \cdot e + (d_1b\{e\})''$. Operating with $a\{e\}$ on the left, we get

$$a\{e\}' = a\{e\} \cdot w + (a\{e\}d_1)''.$$

Hence, by comparison of the components in D'',

$$\tau\{e\} = d_1b\{e\} - a\{e\}d_1 = a\{e\}(-d_1) - (-d_1)b\{e\},$$

which proves Corollary 3.

---

[6] This was proved by H. Cartan; see footnote 4.

It is easy to verify that the additive endomorphism $x \rightarrow a\{x\}d - db\{x\}$ of D still has the formal properties of τ. Thus, our result is also an extension theorem. If we take $E = D$ and $a = 1 = b$, we conclude in particular that every derivation of D over K is an inner derivation.

THEOREM 4.2. *Let D be of finite dimension over K. Then D is normal over K if and only if every indecomposable K-regular (D,D)-space is of dimension* 1 *over D.*

Proof. The necessity of the condition follows at once from the remark following the proof of Theorem 4.1.

Conversely, suppose that the condition of our theorem is satisfied. Consider the Kronecker product $D \times D$ with respect to K. As a vector space over D this is of finite dimension $[D:K]$. Hence $D \times D$ can be decomposed into the direct sum of a finite number of indecomposable (D,D)-subspaces. Say $D \times D = M_1 + \cdots + M_s$. In particular, $1 \times 1 = m_1 + \cdots + m_s$, with $m_i \varepsilon M_i$. We have then $D \cdot m_i \cdot D = M_i$, and $k \cdot m_i = m_i \cdot k$ for each i, and every $k \varepsilon K$. Thus each M_i is K-regular, and hence has dimension 1 over D, so that $M_i = D \cdot m_i$. Hence, for every $d \varepsilon D$, $m_i \cdot d = g_i\{d\} \cdot m_i$, and it is evident that the mapping g_i of D into itself which is thus defined is an automorphism of D over K. If an element $d \varepsilon D$ is left fixed by every g_i it follows that $(1 \times 1) \cdot d = d \cdot (1 \times 1)$, or $1 \times d = d \times 1$, which implies that $d \varepsilon K$. Hence D is normal over K.

COROLLARY. *If D is finite and normal over K, and $D \supseteq E \supseteq K$, then D is normal over E.*

Proof. Every indecomposable E-regular (D,D)-space is a fortiori K-regular, and hence of dimension 1 over D. Hence D is normal over E.

We have seen above that if D is finite and normal over K the classes of (D,D)-operator isomorphic simple K-regular (D,D)-spaces are in a one to one correspondence with the elements of the group of outer automorphisms of D over K (i. e., the full automorphism group, reduced modulo the inner automorphisms). In this correspondence, the Kronecker product of two simple (D,D)-spaces in the classes determined by the outer automorphisms α and β, respectively, belongs to the class determined by $\alpha\beta$. Hence:

THEOREM 4.3. *The classes of operator isomorphic simple K-regular (D,D)-spaces, under the operation induced by Kronecker multiplication of (D,D)-spaces, constitute a group which is isomorphic with the group of outer automorphisms of D over K.*

5. Relative cyclic spaces.

Definition 5.1. Let D be a division ring, E an arbitrary ring with an identity element. Let M be a (D, E)-space which is generated by a single element $m \varepsilon M$, i. e., $M = D \cdot m \cdot E$. Then the pair (M, m) will be called a relative cyclic (D, E)-space.

Definition 5.2. Let (M, m) and (N, n) be two relative cyclic (D, E)-spaces. We shall say that (N, n) covers, or is a cover of, (M, m) if there is a (D, E)-operator homomorphism of N onto M which maps n into m. If (N, n) and (M, m) cover each other, i. e., if they are operator isomorphic in such a way that n and m correspond to each other, they are said to be equivalent.

Let V be an arbitrary (D, E)-space, S an arbitrary subset of V. By a (D, E)-mapping of S into V we shall mean a mapping ϕ such that there exist elements $d_1, \cdots, d_r$ in D and $e_1, \cdots, e_r$ in E with $\phi\{s\} = \sum_{i=1}^{n} d_i \cdot s \cdot e_i$, for every $s \varepsilon S$. If, for $d \varepsilon D$ and $e \varepsilon E$, we define $(d \cdot \phi)\{s\} = d \cdot \phi\{s\}$, and $(\phi \cdot e)\{s\} = \phi\{s\} \cdot e$, we obtain in the additive group of all (D, E)-mappings of S into V the structure of a (D, E)-space. The identity mapping i of S into V is evidently a (D, E)-mapping, and $D \cdot i \cdot E$ coincides with the set of all (D, E)-mappings of S into V.

The set of all (D, E)-mappings of V into V has the natural structure of a ring, H say. On the other hand, if h_0 is the identity mapping of V into V, the pair (H, h_0) is a relative cyclic (D, E)-space, according to what we have just seen. It is easy to characterize these spaces: In fact, if h_1 is an arbitrary element of H we may form the relative cyclic (D, E)-space $(D \cdot h_1 \cdot E, h_1)$. For $h \varepsilon H$, define $\omega\{h\} = hh_1$. Then ω is clearly a (D, E)-operator homomorphism of H onto $D \cdot h_1 \cdot E$ $(= Hh_1)$ and $\omega\{h_0\} = h_1$. Hence (H, h_0) covers $(D \cdot h_1 \cdot E, h_1)$.

Conversely, suppose that (M, m_0) is a relative cyclic (D, E)-space such that, for every $m_1 \varepsilon M$, (M, m_0) covers $(D \cdot m_1 \cdot E, m_1)$. Let h_0 denote the identity mapping of M onto M, H the (D, E)-space consisting of all (D, E)-mappings of M into M. Then we claim that (M, m_0) is equivalent to (H, h_0). Evidently, (H, h_0) covers (M, m_0) by the operator homomorphism ψ, where $\psi\{h\} = h\{m_0\}$. Moreover, if $h\{m_0\} = 0$ then, since (M, m_0) covers $(D \cdot m_1 \cdot E, m_1)$, we have also $h\{m_1\} = 0$, for every $m_1 \varepsilon M$. Hence $h\{m_0\} = 0$ implies that $h = 0$, i. e., ψ is an isomorphism, which proves our assertion.

Thus, the full operator rings of arbitrary (D, E)-spaces may be regarded as special examples of relative cyclic (D, E)-spaces. In the case where D and E are commutative, it is seen at once from the above characterization that every relative cyclic (D, E)-space is equivalent to its own operator ring.

The study of relative cyclic (D, E)-spaces is greatly facilitated by the fact that there is a one to one correspondence between the equivalence classes of finite dimensional relative cyclic (D, E)-spaces and the finite dimensional (D^*, E^*)-subspaces of the space of all additive homomorphisms of E into D.

Let (M, m_0) be a relative cyclic (D, E)-space, and denote by $S(M)$ the set of all D-linear mappings of M into D, regarding M and D as vector spaces over D in the natural fashion. We shall regard $S(M)$ as a (D^*, E^*)-space as follows: If $\sigma \varepsilon S(M)$, $d \varepsilon D$, and $e \varepsilon E$, we define $d^* \cdot \sigma$ as the composite mapping $d^*\sigma$ (σ followed by d^*), and $\sigma \cdot e^*$ by setting, for $m \varepsilon M$, $(\sigma \cdot e^*)\{m\} = \sigma\{m \cdot e\}$.

Now with each $\sigma \varepsilon S(M)$ we associate an additive homomorphism σ' of E into D, where $\sigma'\{e\} = \sigma\{m_0 \cdot e\}$. Then $(d^* \cdot \sigma)' = d^*\sigma'$, and $(\sigma \cdot e^*)' = \sigma' e^*$, so that the σ' for $\sigma \varepsilon S(M)$ constitute a (D^*, E^*)-subspace $R(M, m_0)$ of the space of all additive homomorphisms of E into D. The mapping $\sigma \rightarrow \sigma'$ is a (D^*, E^*)-operator isomorphism of $S(\mathrm{M})$ onto $R(M, m_0)$. We shall call $R(M, m_0)$ the relations space of (M, m_0).

THEOREM 5.1. *If either M is of finite dimension $[M:D]$ over D, or $R(M, m_0)$ is of finite dimension $[R(M, m_0):D^*]$ over D^* then $[M:D] = [R(M, m_0):D^*]$.*

Proof. We can find a set (e_α) of elements of E such that the elements $m_0 \cdot e_\alpha$ constitute a basis for M over D. Let σ_β be the D-linear mapping of M into D which is such that $\sigma_\beta\{m_0 \cdot e_\alpha\} = \delta_{\beta\alpha}$. If $[M:D]$ is finite and $\sigma \varepsilon S(M)$ we have evidently $\sigma = \sum_\alpha (\sigma\{m_0 \cdot e_\alpha\})^* \sigma_\alpha$. Since the σ_α are linearly independent over D^*, it follows that $[S(M):D^*] = [M:D]$. Since the mapping $\sigma \rightarrow \sigma'$ is a D^*-linear isomorphism, this gives $[R(M, m_0):D^*] = [M:D]$. If M is not finite over D, the σ'_α are infinitely many elements of $R(M, m_0)$ which are linearly independent over D^*, so that $R(M, m_0)$ cannot be finite over D^*. This completes the proof.

THEOREM 5.2. *(M, m_0) covers (N, n_0) if and only if $R(M, m_0) \supseteq R(N, n_0)$.*

Proof. Suppose that (M, m_0) covers (N, n_0) by the operator homomorphism h. Then, if $\tau \varepsilon S(\mathrm{N})$, we have $\tau h \varepsilon S(M)$, and $(\tau h)' = \tau'$ which shows that $R(M, m_0) \supseteq R(N, n_0)$.

Conversely, suppose that $R(M, m_0) \supseteq R(N, n_0)$. Then if $d_a \varepsilon D$ and $e_a \varepsilon E$ are such that $\sum_a d_a \cdot m_0 \cdot e_a = 0$ we have

$$\sum_a d_a \sigma'\{e_a\} = \sum_a d_a \sigma\{m_0 \cdot e_a\} = \sigma\{\sum_a d_a \cdot m_0 \cdot e_a\} = 0,$$

for all $\sigma' \varepsilon R(M, m_0)$. A fortiori, $\sum_a d_a \tau'\{e_a\} = 0$, for all $\tau' \varepsilon R(N, n_0)$, i.e., $\tau\{\sum_a d_a \cdot n_0 \cdot e_a\} = 0$, for all $\tau \varepsilon S(N)$. But this clearly implies that $\sum_a d_a \cdot n_0 \cdot e_a = 0$. Hence there is a mapping h of M onto N such that $h\{\sum_a d_a \cdot m_0 \cdot e_a\} = \sum_a d_a \cdot n_0 \cdot e_a$, for all $d_a \varepsilon D$ and $e_a \varepsilon E$. Thus, (M, m_0) covers (N, n_0), and our proof is complete.

Theorem 5.2 implies that a relative cyclic space is determined to within an equivalence by its relations space. Thus, for instance, a necessary and sufficient condition for the relative cyclic (D, E)-space (M, m_0) to be equivalent to the full operator ring of some (D, E)-space (cf. above) is that $D'R(M, m_0)E' = R(M, m_0)$, where D' denotes the ring of left multiplications $x \rightarrow d'\{x\} = dx$ effected by the elements of D in D, and similarly E' denotes the ring of left multiplications in E.

The following theorem establishes the one-to-one correspondence between relative cyclic spaces and spaces of additive homomorphisms which we announced above.

THEOREM 5.3. *Let T be a (D^*, E^*)-subspace of the space of all additive homomorphisms of E into D such that $[T : D^*]$ is finite. Then there exists a relative cyclic (D, E)-space (M, m_0) such that $R(M, m_0) = T$.*

Proof. We may regard D as a vector space over D^*, such that $d_1^* \cdot d = dd_1$. With this understanding, let M be the additive group of all D^*-linear mappings of T into D. We make M into a (D, E)-space by setting, for $d \varepsilon D$, $m \varepsilon M$, $U \varepsilon T$, and $e \varepsilon E$, $(d \cdot m)\{U\} = dm\{U\}$, and $(m \cdot e)\{U\} = m\{Ue^*\}$. Let m_0 be the mapping $U \rightarrow m_0\{U\} = U\{1\}$. Then $m_0 \varepsilon M$. We claim that $D \cdot m_0 \cdot E = M$. In fact, by Lemma 2.1, there exist elements $e_1, \cdots, e_n$ in E and a basis $U_1, \cdots, U_n$ of T over D^* such that $U_i\{e_j\} = \delta_{ij}$. If m is an arbitrary element of M we have

$$(\sum_{i=1}^{n} m\{U_i\} \cdot m_0 \cdot e_i)\{U_j\} = \sum_{i=1}^{n} m\{U_i\} U_j\{e_i\} = m\{U_j\},$$

whence $m = \sum_{i=1}^{n} m\{U_i\} \cdot m_0 \cdot e_i$.

Let U be an arbitrary element of T, and define $\bar{U}\{m\} = m\{U\}$. Then $\bar{U}\{d \cdot m\} = d\bar{U}\{m\}$, i. e., $\bar{U} \varepsilon S(M)$. Furthermore, $\bar{U}'\{e\} = \bar{U}\{m_0 \cdot e\} = U\{e\}$, i. e., $U = \bar{U}' \varepsilon R(M, m_0)$. Thus $T \subseteq R(M, m_0)$. But $[R(M, m_0) : D^*] = [M : D] = [T : D^*]$, whence $T = R(M, m_0)$.

6. Least common covers and decompositions.

Definition 6. 1. Let (M_α, m_α) be any set of relative cyclic (D, E)-spaces. By a least common cover of the (M_α, m_α) we understand a relative cyclic (D, E)-space (P, p_0) which covers each (M_α, m_α) and is covered by every relative cyclic space which covers each (M_α, m_α).

It is clear that, to within equivalence, there is at most one least common cover for a given set of (D, E)-spaces. We shall first prove an existence theorem:

THEOREM 6. 1. *Every set of relative cyclic (D, E)-spaces has a least common cover.*

Proof. Let M be the direct sum of all the M_α. Denote by P the (D, E)-space consisting of all (D, E)-mappings of the set (m_α) into M. If p_0 is the identity mapping of the set (m_α) into M we have $p_0 \varepsilon P$, and $D \cdot p_0 \cdot E = P$. Evidently, the relative cyclic (D, E)-space (P, p_0) covers each (M_α, m_α). On the other hand, if (Q, q_0) covers each (M_α, m_α) then $\sum_\beta d_\beta \cdot q_0 \cdot e_\beta = 0$ implies that $\sum_\beta d_\beta \cdot m_\alpha \cdot e_\beta = 0$, for each α, and hence that $\sum_\beta d_\beta \cdot p_0 \cdot e_\beta = 0$, which shows that (Q, q_0) covers (P, p_0).

THEOREM 6. 2. *Let (M, m_0) and (N, n_0) be relative cyclic (D, E)-spaces of finite dimension over D. Let (P, p_0) be a least common cover of (M, m_0) and (N, n_0). Then $R(P, p_0) = R(M, m_0) + R(N, n_0)$.*

Proof. It is clear from Theorem 5. 2 that $R(P, p_0)$ contains $R(M, m_0) + R(N, n_0)$. On the other hand, by Theorem 5. 3, we can construct a relative cyclic (D, E)-space (Q, q_0) such that $R(Q, q_0) = R(M, m_0) + R(N, n_0)$. Then (Q, q_0) covers (M, m_0) and (N, n_0), and hence also (P, p_0), so that $R(Q, q_0) \supseteq R(P, p_0)$, which gives the desired equality.

Let (M, m_0) be a relative cyclic (D, E)-space, and suppose that we can decompose M into the direct sum of a finite number of (D, E)-subspaces, say $M = M_1 + \cdot \cdot \cdot + M_s$. Then $m_0 = m_1 + \cdot \cdot \cdot + m_s$, with $m \varepsilon M_i$, and $D \cdot m_i \cdot E = M_i$.

Let S_i be the subspace of $S(M)$ which consists of all σ such that

$\sigma\{M_j\} = (0)$, for $j \neq i$. Clearly, $S(M)$ is the direct sum of the S_i, and S_i is isomorphic with $S(M_i)$ in a natural fashion. If $\sigma \varepsilon S(M)$ we write $\sigma = \sigma_1 + \cdots + \sigma_s$, with $\sigma_i \varepsilon S_i$. Then we have $\sigma\{m_0 \cdot e\} = \sigma_1\{m_1 \cdot e\} + \cdots + \sigma_s\{m \cdot e\}$, which shows that each $\sigma' \varepsilon R(M, m_0)$ may be written in the form $\sigma' = \sigma_1' + \cdots + \sigma_s'$, where σ_i' is the element of $R(M_i, m_i)$ which corresponds to σ_i. Moreover, it is clear from what we have said that if there is arbitrarily given, for each index i, an element ρ_i of $R(M_i, m_i)$ we can find $\sigma \varepsilon S(M)$ such that $\sigma_i' = \rho_i$, for each i. Hence $R(M, m_0)$ is equal to the sum of the $R(M_i, m_i)$. Finally, we claim that this sum is direct. In fact, if we have $\sigma_1' + \cdots + \sigma_s' = 0$ we obtain from the above $\sigma\{m_0 \cdot e\} = 0$, for all $e \varepsilon E$, whence $\sigma = 0$, and therefore each $\sigma_i' = 0$.

In the finite dimensional case, we can obtain the following complete result:

THEOREM 6.3. *Let (M, m_0) be a finite dimensional relative cyclic (D, E)-space. Then M is decomposable if and only if $R(M, m_0)$ is decomposable, and the number of indecomposable components is the same for each.*

Proof. There remains only to show that a decomposition of $R(M, m_0)$ leads to a decomposition of M. Suppose that $R(M, m_0)$ is the direct sum of two (D^*, E^*)-subspaces T_1 and T_2. From the proof of Theorem 5.3, we know that (M, m_0) is equivalent with a relative cyclic space (N, n_0), where N is the space of all D^*-linear mappings of $R(M, m_0)$ into D. Clearly, N is the direct sum of two (D, E)-subspaces N_1 and N_2, where N_i is the set of all elements of N which map T_i into (0). This, together with the above, evidently suffices to establish Theorem 6.3.

7. Products.

Definition 7.1. Let C and D be division rings, E an arbitrary ring with an identity element. Let (M, m_0) be a relative cyclic (C, D)-space, (N, n_0) a relative cyclic (D, E)-space. Construct the Kronecker product (C, E)-space $M \times N$ with respect to D, and let L denote the subspace $C \cdot (m_0 \times n_0) \cdot E$. Then the pair $(L, m_0 \times n_0)$ is called the product of (M, m_0) by (N, n_0).

THEOREM 7.1. *Let (M, m_0) and $(N, n\ )$ be as above, and assume that M is finite over C and N is finite over D. Let $(L, m_0 \times n_0)$ be their product. Then $R(L, m_0 \times n_0) = R(M, m_0) R(N, n_0)$.*

Proof. For every $\tau \varepsilon S(N)$ we can define a C-linear mapping $\bar{\tau}$ of

$M \times N$ into M such that $\bar{\tau}\{m \times n\} = m \cdot \tau\{n\}$. Then, for every $\sigma \varepsilon S(M)$, we have $\sigma\bar{\tau} \varepsilon S(M \times N)$. Let ρ be the restriction of $\sigma\bar{\tau}$ to L, and denote by ρ' the corresponding element of $R(L, m_0 \times n_0)$. Then we have clearly $\rho' = \sigma'\tau'$ which shows that $R(M, m_0)R(N, n_0) \subseteq R(L, m_0 \times n_0)$.

Now choose elements $d_1, \cdots, d_r$ in D and $e_1, \cdots, e_s$ in E such that the $m_0 \cdot d_i$ and $n \cdot e_j$ constitute bases for M over C and N over D, respectively. Then the elements $m_0 \cdot d_i \times n_0 \cdot e_j$ constitute a basis for $M \times N$ over C. We have $m_0 \times n_0 \cdot e = \sum_j m_0 \times \tau'_j\{e\} \cdot n_0 \cdot e_j = \sum_j m_0 \cdot \tau'_j\{e\} \times n_0 \cdot e_j = \sum_{i,j} \sigma'_i\tau'_j\{e\} \cdot m_0 \cdot d_i \times n_0 \cdot e_j$, where the $\tau'_j \varepsilon R(N, n_0)$ and the $\sigma'_i \varepsilon R(M, m_0)$.

By Lemma 2.1, we can find elements $f_1, \cdots, f_t$ in E and a basis $\rho'_1, \cdots, \rho'_t$ for $R(M, m_0)R(N, n_0)$ over C^* such that $\rho'_p\{f_q\} = \delta_{pq}$. Then we may write $\sigma'_i\tau'_j = \sum_{r=1}^{t} c^*_{ijp}\rho'_p$, with $c_{ijp} \varepsilon C$, and

$$m_0 \times n_0 \cdot e = \sum_{i,j,p} \rho'_p\{e\}c_{ijp} \cdot m_0 \cdot d_i \times n_0 \cdot e_j.$$

Hence

$$m_0 \times n_0 \cdot f_q = \sum_{i,j} c_{ijq} \cdot m_0 \cdot d_i \times n_0 \cdot e_j,$$

and, on substituting back, $m_0 \times n_0 \cdot e = \sum_p \rho'_p\{e\} \cdot m_0 \times n_0 \cdot f_p$. If $\gamma \varepsilon S(L)$ this gives $\gamma'\{e\} = \sum_p \rho'_p\{e\}\gamma'\{f_p\}$, i. e., $\gamma' = \sum_p \gamma'\{f_p\}^*\rho'_p$, whence $R(L, m_0 \times n_0) \subseteq R(M, m_0)R(N, n_0)$. This completes the proof.

From now on, let E denote a division ring, and D a division subring of E. Then, if (M, m_0) and (N, n_0) are relative cyclic (D, E)-spaces we may form their product with respect to D, as in Definition 7.1, regarding M as a (D, D)-space.

Definition 7.2. The relative cyclic (D, E)-space (M, m_0) is said to be closed if it covers the product of (M, m_0) by (M, m_0).

By Theorems 7.1 and 5.2, we have immediately:

THEOREM 7.2. *If M is finite over D then the relative cyclic (D, E)-space (M, m_0) is closed if and only if $R(M, m_0)$ is a ring.*

Denote by $K(M, m_0)$ the division subring of D which consists of all elements $d \varepsilon D$ for which $d \cdot m_0 = m_0 \cdot d$. It is clear that $K(M, m_0)$ coincides with the set of all elements $d \varepsilon D$ whose left multiplications $e \rightarrow de$ in E commute with all the elements of $R(M, m_0)$. This fact can be used to prove the following result:

THEOREM 7.3. *Suppose that (M, m_0) is closed and finite over D. Then*

$R(M, m_0)$ is the ring of all $K(M, m_0)$-linear mappings of E into D, and $[E : K(M, m_0)] = [M : D]$.

Proof. The first part follows at once from Theorems 7.2 and 2.1, using the above remark. Further, from Theorem 2.1, we have that E is finite over $K(M, m_0)$ and $[E : K(M, m_0)] = [R(M, m_0) : D^*]$, whence the result, by Theorem 5.1.

COROLLARY. *If (M, m_0) and (N, n_0) are closed and finite over D then they are equivalent if and only if $K(M, m_0) = K(N, n_0)$.*

The property of closure characterizes the Kronecker products, as follows:

THEOREM 7.4. *Let E be a division ring, D a division subring of E, and K a division subring of D over which E is finite. Then if $D \times E$ denotes the Kronecker product with respect to K, the relative cyclic (D, E)-space $(D \times E, 1 \times 1)$ is closed. Every closed relative cyclic (D, E)-space (M, m_0) such that M is finite over D is equivalent with a Kronecker product $(D \times E, 1 \times 1)$, where $D \times E$ is taken with respect to $K(M, m_0)$.*

Proof. Evidently, if (N, n_0) is any relative cyclic (D, E)-space such that $K(N, n_0) \supseteq K$, then $(D \times E, 1 \times 1)$, where $D \times E$ is taken with respect to K, is a cover of (N, n_0). Taking for (N, n_0) the product of $(D \times E, 1 \times 1)$ by itself, we see that $(D \times E, 1 \times 1)$ is closed. Since $K(D \times E, 1 \times 1) = K$, the rest now follows from the above Corollary.

HARVARD UNIVERSITY.

EXTENSION TYPES OF ABELIAN GROUPS.*

By REINHOLD BAER.

When studying the manifold of extensions of the group S, we shall employ two principles of classification: equivalence classes and extension types. Here we say that the extensions G and H of S are equivalent extensions of S, if there exists an isomorphy between G and H which leaves invariant every element in S; and the extensions G and H of S belong to the same extension type of S, if there exists a homomorphism of G into H and a homomorphism of H into G which both leave invariant every element in S. It is clear that equivalent extensions belong to the same extension type. If on the other hand G is some extension of S, then direct sums $G \oplus T$ will, in general, not be equivalent extensions of S, though they all belong clearly to the same extension type. It is just this observation which justifies the introduction of extension types; for in constructing extensions of the group S the step from the extension G to the extension $G \oplus T$ will be considered a trivial step. The extension types have another advantage. They form a partially ordered set, if we define: the extension type of the extension G of S is smaller than or equal to the extension type of the extension H of S, if there exists a homomorphism of G into H which leaves invariant every element in S. The first problem then is to give an internal characterization of this partially ordered set of extension types.

It is advantageous to begin such an investigation with the discussion of a somewhat more general problem: given an extension G of some group S and a homomorphism η of S into some group H; to find conditions necessary and sufficient for the existence of a homomorphism of G into H which induces η in S. [If η happens to be an isomorphism, as it will be in our applications, then one may ask the deeper question under which circumstances η is induced by an isomorphism of G upon H.] To answer such questions one has to find invariants of the extension G of S; and these invariants can be used to characterize equivalence classes and the partially ordered set of extension types.

Little is known as yet concerning the problems which we have outlined. Thus we shall give here a complete and concrete solution of them for a restricted class of extensions only. We term the abelian group G a "little"

---

* Received September 5, 1947; revised September 28, 1948.

extension of its subgroup S, if G/S does not contain elements of order 0, and if the presence of an infinity of elements of order a prime p in G implies the absence of elements of order p in G/S. The partially ordered set of types of little extensions of S may then be characterized by means of systems of Loewy chains of S [3, Characterization Theorem]. One may show furthermore that the little extensions G and H of S belong to the same type if, and only if, G and H possess direct summands which are equivalent extensions of S; and that G and H are equivalent extensions if, and only if, they belong to the same type and are isomorphic [3, Theorem 1 and 1, Equivalence Criterion].

The restriction to little extensions has been dictated mainly by the method of proof used. Further restriction, say to finite groups, would have gained little in simplicity, since in particular the rather intricate proof of the existence of extensions with given invariants [in section 2] derives its main complications from a situation which arises already with finite groups.

There appears to be little connection between what is usually termed a group extension and the study of extensions as proposed here. This will be quite clear, if one just inspects the definitions: an extension of the group S is any group which contains S as a subgroup; a group extension, however, is a pair (G, η) where η is a homomorphism of the group G and G is then termed an extension of the kernel of η by the image of G under η.

Notations.

The groups considered will be abelian groups throughout; and their composition will be addition $x + y$. The order of the group element x is 0, if $nx = 0$ implies $n = 0$; and otherwise the order of x is the smallest positive integer n such that $nx = 0$.

$K[p, X] =$ set of all the elements x in the group X such that $px = 0$.

If z is an element in, and T a subset of, the group X, then $\{z\}$ and $\{T\}$ are the subgroups of X which are generated by z and T respectively.

$V \cap W =$ cross cut of the sets V and W.

$V \oplus W =$ direct sum of groups V and W.

$V \simeq W$ signifies isomorphy of groups V and W.

The group P is of type p^∞ [in the sense of Prüfer], if it is generated by elements $g(i)$ subject to the relations: $pg(0) = 0$, $pg(i+1) = g(i)$. Here, as always, p indicates a prime number.

Since all the groups considered are abelian, sums and differences of their homomorphisms and endomorphisms may be defined in the customary fashion.

1. Extension of homomorphisms and isomorphisms. If S is a subgroup of the [commutative] group G, then we say alternately that *G is an extension of S.* If furthermore σ and γ are homomorphisms of S and G respectively such that $x\sigma = x\gamma$ for every element x in S, then we say that *σ is induced by γ in S* and that *γ is an extension of σ.*

PROPOSITION 1. *Suppose that the subgroup S of the group G and the group H satisfy the following conditions:*

(a) *No element in G/S has order* 0.

(b) *If G/S contains elements of order [the prime] p, then H contains only a finite number of elements of order p.*

Then the homomorphism σ of S into H is induced by a homomorphism of G into H if, and only if,

(*) $\quad (S \cap p^iG)\sigma \leq p^iH$ *for every prime power* p^i.

Proof. The necessity of condition (*) is an immediate consequence of the fact that homomorphisms of G into H map k-folds of elements in G upon k-folds of elements in H.

Assume now the validity of condition (*). For the purposes of the present proof it will be convenient to say that the group F between S and G is a finite extension of S, if F/S is a finitely generated group. Because of (a) this is equivalent to the assumption that F/S is a finite group. The system Φ of all the finite extensions F of S [which are contained in G] is a partially ordered set with respect to the containedness relation. If F is in Φ, then we denote by θ_F the set of all the homomorphisms of F into H which induce σ in S. We prove first

(i) $\quad \theta_F$ *is, for F in Φ, finite, but not vacuous.*

Since F/S is a finite group, it is the direct sum of cyclic groups of prime power order:

$$F/S = \{S + z_1\} \oplus \cdots \oplus \{S + z_n\}$$

where $S + z_i$ is an element of prime power order m_i in F/S. Then m_iz_i belongs to S and consequently to $S \cap m_iG$. Hence $(m_iz_i)\sigma$ belongs, by Condition (*), to $(S \cap m_iG)\sigma \leq m_iH$; and this implies the existence of an element t_i in H such that $(m_i z_i)\sigma = m_i t_i$. One verifies now without difficulty the existence of one and only one homomorphism ϕ of F into H which induces σ in S and satisfies $z_i\phi = t_i$, since $F = \{S, z_1, \cdots, z_n\}$ and since F/S is the direct sum of the cyclic groups $\{S + z_i\}$. Thus we have shown the existence of at least one homomorphism ϕ in θ_F.

If F/S contains elements of order a prime p, then it follows from Condition (b) that H contains only a finite number of elements of order p. If m is a power of p, then this implies that the equation $mx = 0$ has only a finite number of solutions x in H. This we may apply on each of the $m_i \neq 1$ so that each of the equations $m_i x = 0$ has only a finite number of solutions x in H.

If ϕ and γ are homomorphisms of F into H which both induce σ in S, then $(m_i z_i)\phi = (m_i z_i)\gamma$, since $m_i z_i$ belongs to S. Hence $m_i z_i(\phi - \gamma) = 0$ for every i. It follows from the preceding paragraph of the proof that there exists only a finite number of elements in H which may be images of z_i under the homomorphism $\phi - \gamma$ of F into H, namely the finitely many solutions of the equation $m_i x = 0$ in H. Since $x(\phi - \gamma) = 0$ for every x in S, one sees that the number of different homomorphisms $\phi - \gamma$ with ϕ, γ in θ_F must be finite [there exists only a finite number of homomorphisms of F into H which map S upon 0]. But then θ_F itself is finite; and this completes the proof of the property (i).

(ii) *If F' and F'' are in Φ, and if $F' \leq F''$, then every homomorphism η in $\theta_{F''}$ induces a homomorphism η' in $\theta_{F'}$.*

This is an obvious consequence of our definitions. The mapping of η onto η', described in (ii), is the *projection of $\theta_{F''}$ into $\theta_{F'}$.*

(iii) *If U and V are in Φ, then $U + V$ is in Φ.*

This is obvious.

(iv) *If $F \leq F' \leq F''$ are in Φ, then the projection of $\theta_{F''}$ into θ_F is the product of the projection of $\theta_{F''}$ into $\theta_{F'}$ by the projection of $\theta_{F'}$ into θ_F.*

This is an obvious consequence of the definition of projection as given in (ii).

(v) *There exists at least one single valued function $\eta(F)$ with the properties:*

(v. a) *$\eta(F)$ is in θ_F for every F in Φ;*

(v. b) *If $F' \leq F''$ are in Φ, then $\eta(F')$ is induced by $\eta(F'')$.*

Kurosh (2) has named such a function $\eta(F)$ a complete projection set of our inverse system. The existence of such a complete projection set may either be deduced from the results of Kurosh (1), (2) or from Steenrod (1), Lemma 2.1, p. 665 [since finite sets are bicompact spaces].

If $\eta(F)$ is some function with the properties (v.a) and (v.b), then there exists one and only one homomorphism η of G into H which induces $\eta(F)$ in F for every F in Φ, since every finite subset of G is contained in at least one F in Φ. But this homomorphism η of G into H induces σ in S, since this is done by every $\eta(F)$. This completes the proof.

In the set $(G \rightarrow H; \sigma)$ of all the homomorphisms of G into H which induce σ in S a *topology* may be introduced in the customary way: if η is a homomorphism of G into H which induces σ in S, and if $a(1), \cdots, a(n)$ are a finite number of elements in G, then a neighborhood [the $a(1)$-, $\cdots$, $a(n)$-neighborhood] of η is formed by all the homomorphisms κ in $(G \rightarrow H; \sigma)$ which satisfy $a(i)\eta = a(i)\kappa$ for $i = 1, \cdots, n$.

It follows from the last step of the proof of Proposition 1 that the space $(G \rightarrow H; \sigma)$ is essentially the same as the space of all the functions $\eta(F)$ with Properties (v.a) and (v.b). If we apply now Steenrod (1), Theorem 2.1, p. 666 [whose applicability is assured since finite sets are bicompact spaces] it follows that the space of functions $\eta(F)$ with Properties (v.a) and (v.b) is a bicompact and non-vacuous space. Since this space is essentially the same as the space $(G \rightarrow H; \sigma)$, we have shown the following fact.

COROLLARY 1. *If σ is a homomorphism of the subgroup S of the group G into the group H such that*

(a) *no element in G/S has order* 0,

(b) *the existence of elements of prime number order p in G/S implies that there exists at most a finite number of elements of order p in H,*

(c) $(S \cap p^iG)\sigma \leq p^iH$ *for every prime power* p^i,

then the space $(G \rightarrow H; \sigma)$ of all the homomorphisms of G into H which induce σ in S is bicompact and non-vacuous.

In order to be sure of the validity of conditions (a), (b) of Proposition 1 and Corollary 1 we introduce the following concept.

DEFINITION 1. *The extension G of its subgroup S is a little extension of S, if*

(a) *G/S does not contain elements of order* 0 *and*

(b) *the existence of elements of prime number order p in G/S implies that there exists at most a finite number of elements of order p in G.*

If G is a little extension of its subgroup S, and if K is between G and S, then K is a little extension of S too.

If G is a little extension of its subgroup S, and if G contains an infinity of elements of order p, then every element of order p in G belongs to S; and x belongs to S whenever px belongs to S, so that $S \cap pG = pS$. This very useful property may be used to prove that little extensions of little extensions are little extensions.

Finally we note that the following fact is easily verified. The group G is a little extension of each of its subgroups if, and only if, G contains only a finite number of elements of every given order.

COROLLARY 2. *S is a direct summand of its little extension G if, and only if,*

(†) $p^iS = S \cap p^iG$ *for every prime power* p^i.

Proof. We note first that $p^iS \leq S \cap p^iG$ is always true. Hence Condition (†) is equivalent to the condition

(††) $S \cap p^iG \leq p^iS$ for every prime power p^i.

It is well known that S is a direct summand of G if, and only if, there exists a homomorphism of G upon S which leaves invariant every element in S. Thus S is a direct summand of G if, and only if, the identity mapping of S may be extended to a homomorphism of G into S. Since G is a little extension of S, we may apply Proposition 1 to this extension problem; and it follows from Proposition 1 that the identity mapping of S is induced by a homomorphism of G into S if, and only if, (††) is true. This shows the validity of our corollary.

Example 1. If G is some commutative group, then we may form the subgroup $P(G)$ of all the elements in G whose order is not 0. Then every element, not 0, in $G/P(G)$ has order 0; and so Condition (†) of Corollary 1 is satisfied by the subgroup $P(G)$. But it is well known that there exist groups G such that $P(G)$ is not a direct summand of G. Consequently it is impossible to omit Condition (a) from Definition 1 without invalidating Corollary 2.

Example 2. Suppose that $\{z_i\}$ is a cyclic group of order p^i and that G is the direct sum of the cyclic groups $\{z_i\}$. Denote by S the subgroup of G which is generated by the elements $z_i - pz_{i+1}$. It is easy to see that these elements form a basis of S, that G/S is a group of type p^∞ [in the sense of Prüfer] and that Condition (†) of Corollary 1 is satisfied by the subgroup S of G. But S cannot be a direct summand of G, since G does not contain subgroups of type p^∞. Consequently it is impossible to omit Condition (b) from Definition 1 without invalidating Corollary 2.

Since Corollary 2 is essentially a special case of Proposition 1, one realizes that Examples 1 and 2 show the impossibility of omitting in Proposition 1 the conditions imposed on S, G and H.

The following "splitting criterion" will prove useful on various occasions.

LEMMA 1. *If G is a little extension of its subgroup S, if the endomorphism τ of G leaves invariant every element in S, then there exists a direct decomposition $G = G' \oplus G''$ such that $S \leq G'$, τ induces an automorphism in G' and such that every element in G'' is annihilated by a suitable power of τ.*

Proof. We begin by proving:

(a) If z is an element in G, then the set $z, z\tau, \cdots, z\tau^i, \cdots$ of elements in G is finite.

This is quite obvious, if z is in S, since then this set is a one element set. If z is not in S, then the order of z modulo S is an integer $n > 1$. If p is a prime divisor of n, then G/S contains elements of order p; and the little extension G of S contains therefore only a finite number of elements of order p. Since G contains, for every prime divisor p of n, only a finite number of elements of order p, it follows that G contains only a finite number of solutions of the equation $ny = 0$. Since nz belongs to S, we have

$$n(z - z\tau^i) = nz - (nz)\tau^i = 0,$$

since elements in S are left invariant by every power of the endomorphism τ. Since G contains only a finite number of solutions of the equation $ny = 0$, it follows that $z - z\tau^i$ can assume only a finite number of different values in G; and this completes the proof of (a).

An endomorphism τ with the property (a) we have termed elsewhere [Baer (1), p. 512] almost periodical; and we have shown [Baer (1), Theorem B, p. 512] that almost periodical endomorphisms "split." In the case of abelian groups this implies the existence of a direct decomposition $G = G' \oplus G''$ with the following properties: τ induces an automorphism in

G', and to every element x in G'' there exists a positive integer $k = k(x)$ such that $x\tau^k = 0$. If u is an element in S, then $u = u' + u''$ where u' belongs to G' and u'' belongs to G''. Since u is left invariant by every power of τ, and since u' is mapped upon an element in G' by every power of τ, we find that $u = u\tau^{k(u'')} = u'\tau^{k(u'')} + u''\tau^{k(u'')} = u'\tau^{k(u'')}$ is an element in G'. Hence $S \leq G'$; and this completes the proof.

PROPOSITION 2. *If G and H are little extensions of their subgroups S and T respectively, then the following properties of the isomorphism σ of S upon T are equivalent:*

(i) *$(S \cap p^iG)\sigma = T \cap p^iH$ for every prime power p^i.*

(ii) *σ is induced in S by a homomorphism of G into H and σ^{-1} is induced in T by a homomorphism of H into G.*

(iii) *There exist direct decompositions $G = G' \oplus G''$ and $H = H' \oplus H''$ such that $S \leqq G'$, $T \leqq H'$ and such that σ is induced in S by an isomorphism of G' upon H'.*

Proof. We show first:

(a) If G *or* H contains an infinity of elements of order p, then neither G/S nor H/T contains elements of order p.

Assume, for instance, that G contains an infinity of elements of order p. Since G is a little extension of S, this implies that G/S does not contain elements of order p. Consequently all the infinitely many elements of order p in G belong to S. But S and T are isomorphic; and so T [and hence H] contains an infinity of elements of order p. But H is a little extension of T; and consequently H/T does not contain elements of order p. Hence (a) is true.

If (i) is satisfied by σ, then $(S \cap p^iG)\sigma \leq p^iH$ and $(T \cap p^iH)\sigma^{-1} \leq p^iG$ for every prime power p^i. Hence we may infer from (a) and Proposition 1 [noting that G/S and H/T are free of elements of order 0] the existence of a homomorphism of G into H which induces σ in S and the existence of a homomorphism of H into G which induces σ^{-1} in T. Thus (ii) is a consequence of (i).

Assume next the validity of (ii). Then there exists a homomorphism γ of G into H which induces σ in S and a homomorphism η of H into G which induces σ^{-1} in T. Then $\tau = \gamma\eta$ is an endomorphism of G which leaves invariant every element in S. We infer from Lemma 1 the existence of a direct decomposition

$$G = G' \oplus G'', \qquad S \leq G'$$

such that τ induces an automorphism in G' and such that every element in G'' is annihilated by a suitable power of τ. We let $H' = G'\gamma$ and we denote by H'' the set of all the elements in H which are annihilated by some power of $\eta\gamma$. It is easy to see that H' and H'' are subgroups of H. Furthermore $T = S\sigma = S\gamma \leq G'\gamma = H'$. If x is an element in G' such that $x\gamma = 0$, then $0 = x\gamma\eta$. But $\gamma\eta$ induces an automorphism in G'; and hence $x = 0$. Thus we have shown that γ induces an isomorphism of G' upon H' and that this isomorphism induces σ in S. Suppose now that y is an element in $H' \cap H''$. Then there exists an element z in G' such that $z\gamma = y$ and there exists a positive integer n such that $y(\eta\gamma)^n = 0$. Hence $z(\gamma\eta)^{n+1} = y(\eta\gamma)^n\eta = 0$. But $\gamma\eta$ induces an automorphism in G' so that $z = 0$. Hence $y = z\gamma = 0$ so that $H' \cap H'' = 0$. If r is an element in H, then $r\eta$ is an element in G. Hence $r\eta = g' + g''$ with g' in G' and g'' in G''. Since $\gamma\eta$ induces an automorphism in G', there exists an element b in G' such that $g' = b\gamma\eta$. It is clear that $b\gamma = r'$ belongs to H'. Let $r'' = r - r'$. Since g'' belongs to G'', there exists a positive integer m such that $g''(\gamma\eta)^m = 0$. Hence

$$\begin{aligned} r''(\eta\gamma)^{m+1} &= (r\eta - r'\eta)(\gamma\eta)^m\gamma = (g' + g'' - b\gamma\eta)(\gamma\eta)^m\gamma \\ &= g''(\gamma\eta)^m\gamma = 0, \end{aligned}$$

so that r'' belongs to H''. Consequently r belongs to $H' + H''$; and this completes the proof of the fact that $H = H' \oplus H''$. Thus we have shown that (iii) is a consequence of (ii).

If finally $G = G' \oplus G''$, $H = H' \oplus H''$, $S \leq G'$, $T \leq H'$, and if the isomorphism α of G' upon H' induces σ in S, then we have:

$$(S \cap p^iG)\sigma = (S \cap p^iG')\sigma = (S \cap p^iG')\alpha = T \cap p^iH' = T \cap p^iH$$

for every prime power p^i. Hence (i) is a consequence of (iii); and this completes the proof.

In order to improve upon the preceding result we need the following fact.

LEMMA 2. *If G and H are isomorphic little extensions of their subgroups S and T respectively, if $G = G' \oplus G''$, $S \leq G'$ and $H = H' \oplus H''$, $T \leq H'$, then the isomorphy of G' and H' implies the isomorphy of G'' and H''.*

Proof. Since G/S does not contain elements of order 0, and since $G'' \simeq G/G' \simeq (G/S)/(G'/S)$, it follows that G'' [and likewise H''] does not contain elements of order 0. If G'' contains elements of order p, then G/S [$\simeq (G'/S) \oplus G''$] contains elements of order p too; and this implies that G contains only a finite number of elements of order p. Likewise we see that H contains only a finite number of elements of order p, if H'' contains elements of order p.

If X is any group, then denote by X_p the set of all the elements of order a power of the prime p. This p-component [or primary component] X_p is a subgroup of X; and the absence of elements of order 0 in X implies that X is the direct sum of its primary components.

If G'' *or* H'' contains elements of order p, then we infer from the isomorphy of G and H that G *and* H both contain only a finite number of elements of order p. Thus $G_p = G'_p$ and $H_p = H'_p$ whenever G and H contain an infinity of elements of order p; and in this case both G''_p and H''_p are 0.

Assume now that G and H contain only a finite number of elements of order p. Then $G_p = G'_p \oplus G''_p$ and $H_p = H'_p \oplus H''_p$. From the isomorphy of G and H we infer the isomorphy of G_p and H_p; and from the isomorphy of G' and H' we infer the isomorphy of G'_p and H'_p. But G_p and H_p contain only a finite number of elements of order p; and thus they are direct sums of a finite number of essentially uniquely determined p-groups of rank 1 [cyclic or of type p^∞]. Hence the isomorphies $G'_p \simeq H'_p$ and $G'_p \oplus G''_p \simeq H'_p \oplus H''_p$ imply the isomorphy of G''_p and H''_p. Thus we have shown that corresponding primary components of G'' and H'' are isomorphic; and this implies the desired isomorphy of G'' and H''.

PROPOSITION 3. *If G and H are little extensions of their subgroups S and T respectively, then the isomorphism σ of S upon T is induced by an isomorphism of G upon H if, and only if,*

(a) $(S \cap p^iG)\sigma = T \cap p^iH$ *for every prime power* p^i *and*

(b) $G \simeq H$.

Proof. The necessity of these conditions is almost obvious. If (a) and (b) are valid, then we infer from (a) and Proposition 2 that there exist direct decompositions $G = G' \oplus G''$ and $H = H' \oplus H''$ such that $S \leq G'$, $T \leq H'$ and such that σ is induced by an isomorphism τ' of G' upon H'. It follows from (b) and the preceding Lemma 2 that G'' and H'' are isomorphic; and consequently there exists an isomorphism of G upon H which induces τ' in G' and σ in S.

EQUIVALENCE CRITERION. *The little extensions G and H of their common subgroup S are equivalent extensions of S if, and only if,*

(a) $S \cap p^iG = S \cap p^iH$ *for every prime power* p^i *and*

(b) $G \simeq H$.

Here we term as usual the extensions G and H of S *equivalent*, if there exists an isomorphy of G upon H which leaves invariant every element in S. The criterion is an obvious special case of Proposition 3.

It should be noted that we may substitute in many cases for condition (b) the weaker condition

(b′) $G/S \simeq H/S$;

but that in general this is impossible, as may be seen from easily constructed examples.

Definition 2. *The group G is a minimal extension of its subgroup S, if G is a little extension of S and if $G = G' \oplus G''$ and $S \leq G'$ imply $G'' = 0$ and $G = G'$.*

The importance of this concept stems from the following simple observations.

Corollary 3. *Suppose that G and H are minimal extensions of their subgroups S and T respectively.*

(a) *The isomorphism σ of S upon T is induced by an isomorphism of G upon H if, and only if,*

$$(S \cap p^iG)\sigma = T \cap p^iH \text{ for every prime power } p^i.$$

(b) *The homomorphism η of G into H which satisfies $S\eta = T$ is an isomorphism of G upon H if, and only if, η induces an isomorphism of S upon T such that*

$$(S \cap p^iG)\eta = T \cap p^iH \text{ for every prime power } p^i.$$

Proof. (a) is an immediate consequence of Proposition 2 and Definition 2. The necessity of the conditions of (b) is obvious. If the conditions of (b) are satisfied, then we infer from (a) the existence of an isomorphism τ of H upon G which induces the isomorphism η^{-1} in T. Then $\eta\tau$ is an endomorphism of G which leaves invariant every element in S; and it follows from Lemma 1 and Definition 2 that $\eta\tau$ is an automorphism of G. Consequently η is an isomorphism of G upon H.

Proposition 4. *If G is a little extension of its subgroup S, then there exists one and essentially only one direct decomposition $G = G' \oplus G''$ such that G' is a minimal extension of S.*

Here we term the direct decompositions $G = H' \oplus H'' = K' \oplus K''$ such that $S \leq H'$ and $S \leq K'$ *essentially the same* [*with respect to* S], if there exists an automorphism of G which maps H' upon K', H'' upon K'' and which leaves invariant every element in S.

Proof. Denote by ϵ the identity mapping of S. Then it follows from Corollary 1 that $(G \rightarrow G; \epsilon)$ is a bicompact topological space. Every element in $(G \rightarrow G; \epsilon)$ is an endomorphism of G and some of these are idempotent $[\eta = \eta^2]$; for instance the identity mapping of G. One verifies easily that the set Ξ of all the idempotents in $(G \rightarrow G; \epsilon)$ is a non-vacuous and closed subset of $(G \rightarrow G; \epsilon)$ so that Ξ too is a bicompact topological space.

If the direct summand D of G contains S, then there exists an idempotent endomorphism of G which maps G upon D and leaves invariant every element in D and, therefore, in S. It follows that the cross cut $\Xi \cap (G \rightarrow D; \epsilon)$ is not vacuous—note that $(G \rightarrow D; \epsilon)$ may be considered as a subset of $(G \rightarrow G; \epsilon)$. It follows from Corollary 1 that $(G \rightarrow D; \epsilon)$ is a bicompact topological space. Hence $(G \rightarrow D; \epsilon)$ is a closed subset of $(G \rightarrow G; \epsilon)$ and this implies that

$$\Xi_D = \Xi \cap (G \rightarrow D; \epsilon) \text{ is a non-vacuous closed subset of } (G \rightarrow G; \epsilon)$$

whenever D is a direct summand of G which contains S.

[Note that elements in Ξ_D are idempotent endomorphisms of G which leave invariant every element in S and which leave invariant only elements in D.]

Consider now a well-ordered descending sequence $D(\nu)$ of direct summands of G which all contain S. Then the sets $\Xi_{D(\nu)}$ form a well ordered descending sequence of non-vacuous, closed subsets of the bicompact space $(G \rightarrow G; \epsilon)$. Their cross cut is consequently not vacuous. If κ belongs to the cross cut of all the $\Xi_{D(\nu)}$, then κ is an idempotent endomorphism of G which leaves invariant every element in S and which leaves invariant only elements belonging to the cross cut of all the $D(\nu)$. The image $G\kappa$ is consequently a direct summand of G which contains S and which is contained in the cross cut of all the $D(\nu)$. Consequently there exists a minimal direct summand G' of G which contains S; and it is clear that G' is a minimal extension of S. Since G' is a direct summand of G, we have $G = G' \oplus G''$; and thus we have shown the existence of at least one decomposition $G = G' \oplus G''$ such that G' is a minimal extension of S.

Suppose now that

$$G = H' \oplus H'' = K' \oplus K''$$

are direct decompositions of G with the property that H' and K' are both minimal extensions of S. This implies clearly

$$S \cap p^i H' = S \cap p^i G = S \cap p^i K' \text{ for every prime power } p^i;$$

and now we infer from Corollary 3, (a) the existence of an isomorphism τ' of H' upon K' which leaves invariant every element in S.

Now we infer from Lemma 2 the isomorphy of H'' and K''. Consequently there exists an automorphism of G which induces τ' in G' and maps H'' upon K''; and this automorphism leaves invariant every element in S. Thus we have shown that the direct decompositions $G = H' \oplus H''$ and $G = K' \oplus K''$ are essentially the same; and this completes the proof.

2. The existence of extensions with given invariants. The invariants of extensions which we consider are chains of subgroups of the following type.

DEFINITION 1. *The subgroups* $S(i)$ *of the group* S *form a p-Loewy chain* [*p a prime*], *if* $S(0) = S$ *and* $pS(i) \leq S(i+1) \leq S(i)$ *for* $0 \leq i$.

If G is an extension of S, then the intersections $S \cap p^i G$ form a p-Loewy chain of S. It would easily be possible to define such chains also with transfinite terms; but for our present purposes this is not necessary.

$S(i)/S(i+1)$ is a direct sum of cyclic groups of order p, since $pS(i) \leq S(i+1)$. This justifies the term Loewy chain. One verifies easily that $p^j S(i) \leq S(i+j)$ and that in particular $p^j S \leq S(j)$.

We note finally that the subgroups $p^i S$ for $0 \leq i$ form a p-Loewy chain. We shall refer to this chain as to *the trivial p-Loewy chain of S*.

LEMMA 1. *If* G *is an extension of the group* S, *and if* p *is a prime number, then*

$$[p(S \cap p^i G) \cap p^{i+2} G]/p(S \cap p^{i+1} G)$$
$$\simeq K[p, (S \cap p^i G) + p^{i+1} G]/(K[p, S \cap p^i G] + K[p, p^{i+1} G]) \text{ for } 0 \leq i.$$

Proof. If we let $K = K[p, G]$, then $X \cap K = K[p, X]$ for every subgroup X of G. If we remember this, and if we let in Ore's formula [Ore (1), p. 174, (17)] $A = K$, $B = S \cap p^i G$, $C = p^{i+1} G$, then we find that the group

$$U = K[p, (S \cap p^i G) + p^{i+1} G]/(K[p, S \cap p^i G] + K[p, p^{i+1} G])$$

is isomorphic to the group

$$V = ([K + (S \cap p^i G)] \cap [K + p^{i+1} G]/[K + (S \cap p^{i+1} G)]$$

where we use the fact that $(S \cap p^i G) \cap p^{i+1} G = S \cap p^{i+1} G$.

Consider now the endomorphism of G which maps x upon px. If X is a subgroup of G, then $K + X$ is the inverse image of pX under this endomorphism; and one sees easily that $(K + X) \cap (K + Y)$ is the inverse image of $pX \cap pY$ [where X and Y are subgroups of G]. This implies the isomorphy of the groups V and

$$W = [p(S \cap p^iG) \cap p^{i+2}G]/p(S \cap p^{i+1}G).$$

Thus U and W are isomorphic groups, as we intended to show.

COROLLARY 1. *If G is an extension of the group S, and if p is a prime number, then $[p(S \cap p^iG) \cap (S \cap p^{i+2}G)]/p(S \cap p^{i+1}G)$ is isomorphic to a subgroup of $K[p, p^iG]/K[p, p^{i+1}G]$.*

Proof. It is an immediate consequence of Lemma 1 that the group $[p(S \cap p^iG) \cap (S \cap p^{i+2}G)]/p(S \cap p^{i+1}G)$ is isomorphic to a subgroup of $K[p, p^iG]/(K[p, S \cap p^iG] + K[p, p^{i+1}G])$; and this group is a homomorphic image of $K[p, p^iG]/K[p, p^{i+1}G]$. But every homomorphic image of $K[p, p^iG]/K[p, p^{i+1}G]$ is isomorphic to a subgroup of $K[p, p^iG]/K[p, p^{i+1}G]$ so that the first of our groups is isomorphic to a subgroup of the last one.

EXISTENCE THEOREM. *Suppose that there is given, for every prime p, a p-Loewy chain $S(i;p)$ of the group S. Then there exists a little extension G of S such that $S \cap p^iG = S(i;p)$ for every prime power p^i if, and only if, the following conditions are satisfied by these Loewy chains.*

(N) *If the p-Loewy chain $S(i;p)$ is non-trivial, then*

(a) *S contains only a finite number of elements of order p;*

(b) *$[pS(i;p) \cap S(i+2;p)]/pS(i+1;p)$ is finite for every i; and*

(c) *$[pS(i;p) \cap S(i+2;p)]/pS(i+1;p) = 0$ for almost every i.*

Remark. If the p-Loewy chain $S(i;p)$ is trivial, then conditions (b) and (c) are automatically satisfied, whereas the validity of (a) is not required.

Proof of the necessity of Condition (N). Assume the existence of a little extension G of S such that $S \cap p^iG = S(i;p)$ for every prime power p^i. If G/S does not contain elements of order p, then G/S does not contain elements of order p^i and $S \cap p^iG = p^iS$. Thus the absence of elements of order p in G/S implies the triviality of the p-Loewy chain $S(i;p)$. If the p-Loewy chain $S(i;p)$ is non-trivial, then G/S contains elements of order p. Since G is a little extension of S, this implies the finiteness of $K[p, G]$. From the finiteness of $K[p, G]$ we infer the finiteness of $K[p, S]$, the finiteness of $K[p, p^iG]/K[p, p^{i+1}G]$ for every i and the vanishing of almost every $K[p, p^iG]/K[p, p^{i+1}G]$. The necessity of Condition (N) is now an immediate consequence of Corollary 1.

We precede the proof of the sufficiency of Condition (N) by the proof of a more general existential proposition.

LEMMA 2. *If $S(i)$ is a p-Loewy chain of the group S, then there exists an extension P of S with the following properties:*

(a) $S(i) = S \cap p^iP$ *for every* i;

(b) $K[p; P]/K[p; S]$ *is isomorphic to the direct sum of the groups* $[S(i+2) \cap pS(i)]/pS(i+1)$;

(c) *every element in P/S is of order a power of p.*

Proof. We show by complete induction with respect to i the existence of groups $E(i)$ with the following properties:

(1) $S = E(0)$;

(2) $E(j) \leq E(j+1)$ for $0 \leq j$;

(3) $p^jE(j) = S(j)$;

(4) $p^kE(j) \cap S = S(k)$ for $0 < k < j$;

(5) $K[p, E(1)] = K[p, S]$;

(6) $K[p, E(j+1)]/K[p, E(j)] \simeq [S(j+1) \cap pS(j-1)]/pS(j)$ for $0 < j$.

We note that $S = E(0)$ meets all the relevant requirements. It will be convenient, however, to give a separate construction of $E(1)$. We note first that $pS = pS(0) \leq S(1)$. Hence $S(1)/pS$ is a direct sum of cyclic groups of order p; and consequently there exists some basis B of $S(1)$ modulo pS. We obtain $E(1)$ by adjoining to S elements b', for b in B, subject only to the relations: $pb' = b$.

Then $E(1) = S + \{B'\}$ [where B' indicates the set of all the b' for b in B]. Hence $pE(1) = pS + \{pB'\} = pS + \{B\} = S(1)$.

If x is any element in $E(1)$, then $x = s + \sum c(b)b'$ where s is in S and where almost every $c(b) = 0$ and where the summation ranges over all the b in B. Since $pb' = b$ belongs to S, we may assume without loss in generality that $c(b) = 0$, if divisible by p. If $0 = px$, then $0 = px = ps + \sum c(b)b$. Since ps belongs to pS, and since B is a basis of $S(1)$ modulo pS, it follows that every $c(b)$ is divisible by p. Hence every $c(b) = 0$; and so $x = s$ belongs to S. This completes the proof of the fact that $E(1)$ meets all the relevant requirements.

Assume now that $0 < i$ and that we have constructed, for $0 \leq j \leq i$, groups $E(j)$ which satisfy conditions (1) to (6). Since the $S(k)$ form a p-Loewy chain, it follows from (3) that

$$pS(i) \leq S(i+1) \cap pS(i-1) \leq S(i+1) \leq S(i) = p^iE(i).$$

This implies in particular that $[S(i+1) \cap pS(i-1)]/pS(i)$ and $S(i+1)/[S(i+1) \cap pS(i-1)]$ are direct sums of cyclic groups of order p. Consequently there exists a basis A of $S(i+1)$ modulo $S(i+1) \cap pS(i-1)$ and a basis B of $S(i+1) \cap pS(i-1)$ modulo $pS(i)$. Since A is part of $p^iE(i)$, there exists to every element a in A an element a' in $E(i)$ such that $p^ia' = a$.

We obtain $E(i+1)$ by adjoining to $E(i)$ elements a'', for a in A, and elements b', for b in B, subject solely to the relations:

$$pa'' = a' \text{ for } a \text{ in } A, \; p^{i+1}b' = b \text{ for } b \text{ in } B.$$

We note first that $E(i+1) = E(i) + \{A''\} + \{B'\}$ and that consequently

$$p^{i+1}E(i+1) = pS(i) + \{A\} + \{B\} = [S(i+1) \cap pS(i-1)] + \{A\}$$
$$= S(i+1)$$

since B is a basis of $S(i+1) \cap pS(i-1)$ modulo $pS(i)$, A a basis of $S(i+1)$ modulo $S(i+1) \cap pS(i-1)$, and since $S(i) = p^iE(i)$ by (3).

Every element x in $E(i+1)$ may be represented in the form:

$$\text{(r)} \qquad x = u + \sum_a c(a)a'' + \sum_b c(b)b'$$

where u belongs to $E(i)$, where the first of these sums ranges over all the a in A and the second sum over all the b in B, and where almost every $c(a)$ and almost every $c(b)$ is 0. Since $pa'' = a'$ belongs to $E(i)$, we may assume without loss in generality that $c(a) = 0$ whenever $c(a)$ is divisible by p. Since $p^{i+1}b' = b$ belongs to $S \leq E(i)$, we may assume without loss in generality that $c(b) = 0$ whenever $c(b)$ is divisible by p^{i+1}. We shall refer to representations (r) of x which meet all these requirements as to *normalized* representations.

Suppose now that the element x in $E(i+1)$ is given in the normalized form (r), that $0 < j \leq i$ and that p^jx belongs to S. Then

$$p^jx = p^ju + p^{j-1}\sum_a c(a)a' + \sum_b p^jc(b)b'.$$

Since $p^j x$ is in S, it is in $E(i)$. Since u is in $E(i)$, so is $p^j u$. Furthermore every a' is in $E(i)$. Consequently $\sum_b p^j c(b) b'$ belongs to $E(i)$. It follows from the construction of $E(i+1)$ that every $p^j c(b) b'$ belongs to $E(i)$ and that therefore $p^j c(b)$ is divisible by p^{i+1}. Hence $c(b) = p^{i+1-j} d(b)$ where $d(b) = 0$ whenever $d(b)$ is divisible by p^j. Consequently

$$p^j x = p^j u + p^{j-1} \sum_a c(a) a' + \sum_b d(b) b.$$

Since $p^j x$ as well as all the b belong to S, and since u and all the a' are in $E(i)$, it follows that $p^{j-1}[pu + \sum_a c(a) a']$ belongs to $S \cap p^{j-1}E(i) = S(j-1)$ [by (4)]. Multiplying this element by p^{i-j+1} we infer that $p^{i+1}u + \sum_a c(a) a$ belongs to $p^{i-j+1}S(j-1) \leq pS(i-1)$, since the $S(k)$ form a p-Loewy chain. But $p^{i+1}u$ belongs to $p^{i+1}E(i) = pS(i) \leq pS(i-1)$ by (3); and thus we have shown that $\sum_a c(a) a$ belongs to $pS(i-1)$. Since A is part of $S(i+1)$, $\sum_a c(a) a$ belongs to $S(i+1) \cap pS(i-1)$; and since A forms a basis of $S(i+1)$ modulo $S(i+1) \cap pS(i-1)$, it follows that every $c(a)$ is divisible by p. This implies $c(a) = 0$ because of our normalization of (r). Hence

$$p^j x = p^j u + \sum_b d(b) b.$$

Since $p^j x$ and every b are in S, it follows that $p^j u$ belongs to $S \cap p^j E(i) = S(j)$ by (4). Since every b belongs to $S(i+1) \leq S(j)$ [because of $j < i+1$], we have shown that $p^j x$ belongs to $S(j)$. Now one deduces from (4) easily the inequalities:

$$S(j) = p^j E(i) \cap S \leq p^j E(i+1) \cap S \leq S(j);$$

and this shows the validity of

$$S(j) = p^j E(i+1) \cap S \text{ for } 0 < j < i+1.$$

If b belongs to B, then b belongs to $S(i) = p^i E(i)$. Hence there exist elements b'' in $E(i)$ such that $p^i b'' = b$. From $p^{i+1} b' = b$ we infer that $p[p^i b' - p^{i-1} b''] = 0$; and thus the elements

$$b^* = p^i b' - p^{i-1} b'' \text{ for } b \text{ in } B \text{ belong to } K[p, E(i+1)].$$

Suppose now that $f(b)$, for b in B, are integers such that almost every $f(b) = 0$ and $\sum_b f(b) b^*$ belongs to $K[p, E(i)]$. Since every b'' belongs to $E(i)$, this implies that $\sum_b f(b) p^i b'$ belongs to $E(i)$. It follows from the construction of $E(i+1)$ that every $f(b)p^i b'$ belongs to $E(i)$ and that therefore $f(b)p^i$ is divisible by p^{i+1} and $f(b)$ is divisible by p. Since $pb^* = 0$, this implies $\sum_b f(b)b^* = 0$; and we have shown that

the elements b^* for b in B are independent modulo $K[p, E(i)]$.

Consider now an element x in $K[p, E(i+1)]$; and suppose that it is given in the normalized form (r). Then

$$0 = px = pu + \sum_a c(a)a' + \sum_b pc(b)b'.$$

Since pu and the elements a' are in $E(i)$, it follows that $\sum_b pc(b)b'$ belongs to $E(i)$; and we infer as usual that every $pc(b)$ is divisible by p^{i+1}. Hence $c(b) = p^i g(b)$ where $g(b) = 0$ if divisible by p. Hence

$$0 = pu + \sum_a c(a)a' + \sum_b g(b)b;$$

and multiplying by p^i we find that

$$0 = p^{i+1}u + \sum_a c(a)a + \sum_b p^i g(b)b.$$

The element $p^{i+1}u$ belongs to $p^{i+1}E(i) = pS(i)$ by (3); and the element $p^i b$ belongs to $p^i[S(i+1) \cap pS(i-1)] \leq p[S(i+1) \cap pS(i-1)] \leq pS(i)$ since $0 < i$. Consequently $\sum_a c(a)a$ belongs to $pS(i)$ too. But A is a basis of $S(i+1)$ modulo $S(i+1) \cap pS(i-1)$; and this implies, a fortiori, that A is independent modulo $pS(i)$. Hence every $c(a) = 0$; and now it follows that

$$\begin{aligned} x &= u + \sum_b c(b)b' = u + \sum_b g(b)p^i b' \\ &= u + \sum_b g(b)p^{i-1}b'' + \sum_b g(b)b^* = v + \sum_b g(b)b^*. \end{aligned}$$

Since u and b'' belong to $E(i)$, so does $v = u + \sum_b g(b)p^{i-1}b''$. Since x and b^* belong to $K[p, E(i+1)]$, we have $pv = 0$. Hence v belongs to $K[p, E(i)]$. This completes the proof of the fact that

the elements b^* for b in B form a basis of $K[p, E(i+1)]$ modulo $K[p, E(i)]$.

Thus we have shown that $K[p, E(i+1)]/K[p, E(i)]$ and $[S(i+1) \cap pS(i-1)]/pS(i)$ possess bases containing the same number of elements. This implies the isomorphy of these groups, since they are both direct sums of cyclic groups of order p.

This completes the proof of the fact that the group $E(i+1)$ satisfies all the relevant conditions (1) to (6). Thus we have completed the inductive construction of the groups $E(i)$.

Since the groups $E(i)$ form an ascending chain of groups, there exists one and only one group P which is the sum of the groups $E(i)$. From (1) it follows that $S = E(0)$ is part of P. If x is an element in P such that p^ix belongs to S, then x belongs to some $E(k)$. If $i < k$, then we infer from (4) that p^ix belongs to $p^iE(k) \cap S = S(i)$; and if $k \leq i$, then we infer from (3) that p^ix belongs to $p^iE(k) \cap S = p^{i-k}p^kE(k) \cap S = p^{i-k}S(k) \leq S(i)$, since the $S(j)$ form a p-Loewy chain. Consequently it follows from (3) that

$$S(i) = p^iE(i) \leq p^iP \cap S \leq S(i) \quad \text{or} \quad S(i) = p^iP \cap S.$$

Noting that $K[p;P]$ is the sum of the $K[p;E(i)]$ one sees easily that $K[p,P]$ is isomorphic to the direct sum of the $K[p;E(i+1)]/K[p,E(i)]$; and hence it follows from (5) and (6) that $K[p,P]/K[p,S]$ is isomorphic to the direct sum of the $[S(i+1) \cap pS(i-1)]/pS(i)$ for $0 < i$. If finally x is an element in P, then x belongs to some $E(n)$; and it follows from (3) that p^nx belongs to $p^nE(n) = S(n) \leq S$. Hence every element in P is modulo S of order a power of p. Thus P meets all the requirements; and this completes the proof of Lemma 2.

Proof of the sufficiency of Condition (N). Assume that the p-Loewy chains $S(i;p)$ satisfy Condition (N). If $S(i;p)$, for some prime p, is a trivial p-Loewy chain of S, then we let $S = P_p$. If the p-Loewy chain $S(i;p)$ is, for some prime p, not trivial, then we denote by P_p some extension of S with the properties:

(d) $S(i;p) = S \cap p^iP_p$;

(e) $K[p,P_p]/K[p,S]$ is isomorphic to the direct sum of the groups $[S(i+2;p) \cap pS(i;p)]/pS(i+1;p)$;

(f) every element in P_p/S is of order a power of p.

It follows from (e) and Conditions (N. a) to (N. c) that

(g) $K[p,P_p]$ is finite.

The existence of P_p is a consequence of Lemma 2.

There exists one and essentially only one extension G of S which contains all these groups P_p as subgroups and which is their sum. One deduces from (f) and (g) that G is a little extension of S; and one deduces from (d) and

(f) that $S(i;p) = S \cap p^iG$ for every prime power p^i. Thus G meets all our requirements; and this completes the proof of the Existence Theorem.

We offer two important special cases of the Existence Theorem.

A. *Suppose that S contains only a finite number of elements of any given order; and that $S(i;p)$ is, for every prime number p, a p-Loewy chain of S. Then there exists a little extension G of S such that $G \cap p^iS = S(i;p)$ for every prime power p^i.*

To deduce this from the Existence Theorem one has to note mainly that S does not contain elements of order 0 and that the descending chain condition is satisfied by the subgroups of every given primary component of S. We omit the further details of the derivation.

B. *Suppose that S is a finite group; and that $S(i;p)$ is, for every prime number p, a p-Loewy chain of S. Then there exists a finite extension G of S such that $G \cap p^iS = S(i;p)$ for every prime power p^i if, and only if,*

(B′) *there exists to every prime p an integer $m = m(p)$ such that $S(m(p);p) = 0$ and*

(B″) *there exists only a finite number of primes p with non-trivial p-Loewy chain $S(i;p)$.*

We omit the proof.

3. The partially ordered set of extension types. The following generalization of the concept of equivalence is used for the definition of this structure.

Definition 1. *If G and H are extensions of the group S, and if there exists a homomorphism of G into H which leaves invariant every element in S, then* $\text{Ext}\,[S < G] \leq \text{Ext}\,[S < H]$. *If both* $\text{Ext}\,[S < H] \leqq \text{Ext}\,[S < G]$ *and* $\text{Ext}\,[S < G] \leq \text{Ext}\,[S < H]$, *then* $\text{Ext}\,[S < H] = \text{Ext}\,[S < G]$.

We express this in words by saying that the extension type of H over S is smaller than or equal to the extension type of G over S. It is clear now that the extension types over S form a partially ordered set. We shall restrict ourselves to a consideration of the little extensions of S; and their extension types form a partially ordered set too which we denote by $\Lambda(S)$, *the partially ordered set of the types of little extensions over S.*

To obtain an internal characterization of $\Lambda(S)$ we define another partially ordered set. If $L(p)$ is, for every prime number p, a definite p-Loewy chain of the group S, then we term

$$L = [L(2), L(3), \cdots, L(p), \cdots]$$

a Loewy system of S whose p-th coordinate is the p-Loewy chain $L(p)$. If the p-Loewy chain $L(p)$ consists of the subgroups $S(i;p)$ of S and the p-Loewy chain $L'(p)$ consists of the subgroups $S'(i;p)$ of S, and if $S(i;p) \leq S'(i;p)$, then we let $L(p) \leq L'(p)$. If $L = [\cdots L(p) \cdots]$ and $L' = [\cdots L'(p) \cdots]$ are Loewy systems of S such that $L(p) \leq L'(p)$ for every prime number p, then we let $L \leq L'$. It is clear that in this fashion *the system of all the Loewy systems of* S has been made into a partially ordered set. This partially ordered set is, in general, somewhat too big. Hence we define.

DEFINITION 2. *The Loewy system* $L = [\cdots L(p) \cdots]$ *is a little Loewy system, if it satisfies the following condition*:

(N) *If the p-Loewy chain* $L(p)$ *consisting of the subgroups* $L(i;p)$ *of* S *is non-trivial, then*

(a) S *contains only a finite number of elements of order* p;

(b) $[pL(i;p) \cap L(i+2;p)]/pL(i+1;p)$ *is finite for every* i; *and*

(c) $[pL(i;p) \cap L(i+2;p)]/pL(i+1;p) = 0$ *for almost every* i.

The partially ordered set of the little Loewy systems in S we denote by $L(S)$.

Now we may map the little extension G of S upon the Loewy system whose p-th coordinate is the p-Loewy chain $S \cap p^i G$. It is a consequence of the Existence Theorem of Section 2 that this Loewy system is a little Loewy system; and one deduces readily from 1, Proposition 1 and the Existence Theorem of Section 2 that this mapping induces a [natural] isomorphism of the partially ordered set $\Lambda(S)$ upon the partially ordered set $L(S)$. We restate this result as follows.

CHARACTERIZATION THEOREM. *The partially ordered sets* $\Lambda(S)$ *and* $L(S)$ *are essentially the same.*

The division into extension types and the mapping of $\Lambda(S)$ upon $L(S)$ may be brought into sharper focus by the following results.

THEOREM 1. *The following properties of the little extensions* G *and* H *of* S *are equivalent.*

(i) $\operatorname{Ext}[S < G] = \operatorname{Ext}[S < H]$.

(ii) *$S \cap p^i G = S \cap p^i H$ for every prime power p^i.*

(iii) *If $G = G' \oplus G''$ and $H = H' \oplus H''$ are direct decompositions such that G' and H' are minimal extensions of S, then G' and H' are equivalent extensions of S.*

(iv) *There exist direct decompositions $G = A \oplus B$ and $H = C \oplus D$ such that A and C are equivalent extensions of S.*

This is an almost immediate consequence of 1, Proposition 2 and 1, Proposition 4 and 1, Corollary 3.

Combining Theorem 1 and the Equivalence Criterion of Section 1 we find that the little extensions G and H of S belong to the same equivalence class if, and only if, $G \cong H$ and $\mathrm{Ext}\,[S < G] = \mathrm{Ext}\,[S < H]$. Thus the equivalence classes are determined within the extension type by the structure of the extended group.

THEOREM 2. *The following properties of the little extensions G and H of S are equivalent.*

(i) $\mathrm{Ext}\,[S < G] \leq \mathrm{Ext}\,[S < H]$.

(ii) *$S \cap p^i G \leq S \cap p^i H$ for every prime power p^i.*

(iii) *There exists a little extension J of G [and of S] such that* $\mathrm{Ext}\,[S < J] = \mathrm{Ext}\,[S < H]$.

Proof. The equivalence of conditions (i) and (ii) is an immediate consequence of 1, Proposition 1. If (iii) is true, then there exists a homomorphism γ of J into H which leaves invariant every element in S; and hence it follows from $G \leq J$ that

$$S \cap p^i G \leq S \cap p^i J = (S \cap p^i J)\gamma \leq S \cap p^i H \text{ for every prime power } p^i;$$

and thus (ii) is a consequence of (iii).

Assume finally the validity of (ii). It is a consequence of 1, Proposition 4 that $H = H' \oplus H''$ where H' is a minimal extension of S. Clearly $\mathrm{Ext}\,[S < H] = \mathrm{Ext}\,[S < H']$. Since we intend to show the existence of an extension of G which belongs to the same extension type of S as H, we may assume now without loss in generality that H itself is a minimal extension of S.

We form now a group K with the following properties:

(a) $K = G + H, \qquad S = G \cap H.$

One deduces from (a) readily the validity of the following properties.

(a′) Every element in K has the form $g + h$ with g in G and h in H.

(a″) $g + h$ [with g in G and h in H] belongs to S if, and only if, both g and h belong to S.

Next we show

(b) *$S \cap p^iK = S \cap p^iH$ for every prime power p^i.*

It is clear that $S \cap p^iH \leq S \cap p^iK$. If conversely x is an element in $S \cap p^iK$, then there exist elements g and h in G and H respectively such that $x = p^i(g + h) = p^ig + p^ih$. Since p^ig is in G, p^ih in H and x in S, we infer from (a″) that p^ig and p^ih both belong to S. Hence p^ih belongs to $S \cap p^iH$; and it follows from (ii) that p^ig belongs to $S \cap p^iG \leq S \cap p^iH$. Consequently x is in $S \cap p^iH$; and this completes the proof of (b).

(c) *There exists a homomorphism κ of K into H which leaves invariant every element in S.*

Since $K/S = (G/S) \oplus (H/S)$, and since none of the elements in G/S and H/S is of order 0, none of the elements in K/S is of order 0. If K/S contains an element of prime number order p, then G/S or H/S contains an element of order p. If H/S contains an element of order p, then the little extension H of S contains only a finite number of elements of order p. If G/S contains an element of order p, then the little extension G of S contains only a finite number of elements of order p. Hence S likewise contains only a finite number of elements of order p. This implies that H contains only a finite number of elements of order p, since otherwise S would contain all the infinitely many elements of order p in H. Thus we have shown that H contains only a finite number of elements of order p whenever K/S contains elements of order p. Consequently we may apply 1, Proposition 1; and we infer from (b) the existence of a homomorphism κ of K into H which leaves invariant every element in S.

(d) *H is a direct summand of K.*

By (c) there exists a homomorphism κ of K into H which leaves invariant every element in S. Since $H \leq K$, κ is an endomorphism of K and consequently κ induces an endomorphism of H which leaves invariant every element in S. Since H is a minimal extension of S, it follows from 1, Corollary 3, (b) that κ induces an automorphism in H. If W is the kernel of the endomorphisms κ of K, then it is clear that $W \cap H = 0$, as κ induces an automorphism in H. If x is an element in K, then $x\kappa$ belongs to H. But $H = H\kappa$; and consequently there exists an element y in H such that $y\kappa = x\kappa$. Clearly $x - y$ belongs to W; and thus we have shown that $K \doteq H \oplus W$; and this proves (d).

Let $K = H \oplus N$ by (d). Then $N \simeq K/H = (G + H)/H \simeq G/(G \cap H) = G/S$. Thus N and K contain an infinity of elements of order p whenever G/S contains an infinity of elements of order p; and hence K will, in general, not be a little extension of S. Thus K will not be the desired extension J in spite of the fact that K and H belong by (c) to the same extension type of S.

There exists a maximal subgroup M of N such that $M \cap G = 0$. This follows from the customary application of set theoretical principles [Maximum Principle]. We prove next

(e) *Every element in N/M is of order different from* 0, *and to every prime p there exists only a finite number of elements of order p in N/M.*

Since N and G/S are isomorphic, neither N nor N/M can contain elements of order 0. Suppose now that p is a prime number such that N/M contains an infinity of elements of order p. Then N contains elements of order p. Hence G/S contains elements of order p. Hence the little extension G of S contains only a finite number of elements of order p. Consequently $K[p, G]$ and $(M + K[p, G])/M$ are finite groups. Since therefore $[N \cap (M + K[p, G])]/M$ is a finite subgroup of the group $K[p, N/M]$ which is by hypothesis infinite, there exists an element $M + w$ in $K[p, N/M]$ which does not belong to $[N \cap (M + K[p, G])]/M$. Form $M' = M + \{w\}$. Then $M < M' \leq N$. It follows from the maximality of M that $M' \cap G \neq 0$. Consequently there exists an element m in M and an integer i such that $0 \neq m + iw$ belongs to G. Since $M + w$ is in $K[p, N/M]$, pw is in M. Since $M \cap G = 0$, it follows that i is prime to p. Since pw is in M, so is $pm + ipw$. But this element belongs to G too and $M \cap G = 0$. Hence $p(m + iw) = 0$; and $m + iw$ belongs to $K[p, G]$. Consequently $M + w$ is an element in $[N \cap (M + K[p, G])]/M$. This contradicts our choice of $M + w$; and thus we have been led to a contradiction which proves that there exists only a finite number of elements of order p in N/M.

We let $J = K/M$. From $G \cap M = 0$ it follows that G and $(G + M)/M$ are essentially the same; and thus we may consider J as an extension of G. Since $K/G \simeq H/S$ does not contain elements of order 0, neither does $J/G \simeq K/(G + M)$. If J/G contains elements of order p, so does $K/G \simeq H/S$. Hence H contains only a finite number of elements of order p. But

$J = (H \oplus N)/M \simeq H \oplus (N/M)$; and it follows from (e) that J contains only a finite number of elements of order p. Thus we have shown that J is a little extension of G [where we have identified as usual $(G + M)/M$ and G]. From $J \simeq H \oplus (N/M)$ we infer that $\mathrm{Ext}\,[S < J] = \mathrm{Ext}\,[S < H]$; and thus we have shown that J meets all the requirements. Hence (iii) is a consequence of (ii); and this completes the proof.

Remark 1. If G is a group between S and the extension H of S, then clearly $\mathrm{Ext}\,[S < G] \leq \mathrm{Ext}\,[S < H]$. The question arises whether one may conversely infer from $\mathrm{Ext}\,[S < G] \leq \mathrm{Ext}\,[S < H]$ the existence of a group B between S and H such that $\mathrm{Ext}\,[S < G] = \mathrm{Ext}\,[S < B]$. The following example shows that this is not always true. Let $G = \{u\} \oplus \{v\}$ where u and v are elements of order p and p^3 respectively; and let $S = \{u + pv\}$. Then G is readily seen to be a minimal extension of the cyclic group S of order p^2. Let H be a cyclic group of order p^3 which contains S as a subgroup. Then $H = \{w\}$ where $pw = u + pv$. There exists one and only one homomorphism η of G into H such that $u\eta = p^2 w$, $v\eta = (1 - p)w$. Then $(u + pv)\eta = pw = u + pv$ so that η leaves invariant every element in S. Hence $\mathrm{Ext}\,[S < G] < \mathrm{Ext}\,[S < H]$, since equality of these extension types is clearly impossible. But G and H are both minimal extensions of S; and thus it is impossible [by 1, Corollary 3] to find a group between S and H whose extension type is that of G.

Remark 2. *The system of extension types of a little extension.*

Suppose that the group T is a little extension of the group S. Then $\mathrm{Ext}\,[S < T]$ is a well determined element in $\Lambda(S)$; and we denote by $\Lambda(S;T)$ the system of all the extension types θ in $\Lambda(S)$ such that $\mathrm{Ext}\,[S < T] \leq \theta$.

If G is a little extension of T, then G is likewise a little extension of S which satisfies $\mathrm{Ext}\,[S < T] \leq \mathrm{Ext}\,[S < G]$, so that $\mathrm{Ext}\,[S < G]$ belongs to $\Lambda(S;T)$. If G and H are little extensions of T such that $\mathrm{Ext}\,[T < G] = \mathrm{Ext}\,[T < H]$, then one verifies easily the equality $\mathrm{Ext}\,[S < G] = \mathrm{Ext}\,[S < H]$. Consequently we obtain a single valued and order preserving mapping of $\Lambda(T)$ into $\Lambda(S;T)$, if we map $\mathrm{Ext}\,[T < G]$ upon $\mathrm{Ext}\,[S < G]$; and it follows from Theorem 2 that in this fashion $\Lambda(T)$ is mapped upon the whole system $\Lambda(S;T)$.

Simple examples show that this mapping will in general not be a one to one correspondence.

Remark 3. *The existence of common little extensions.*

The system $\Lambda(S)$ contains one and only one maximal element; it is represented by a little extension G of S which satisfies $G = Gp$ for every prime p such that S contains only a finite number of elements of order p [or equivalent: it is characterized by a little Loewy system $S(i;p) = p^iS$ in case S contains an infinity of elements of order p and $S(i;p) = S$ in case S contains only a finite number of elements of order p].

Suppose now that G and H are little extensions of S. From the preceding remark we infer the existence of extension types Φ in $\Lambda(S)$ such that Ext $[S < G] \leq \Phi$ *and* Ext $[S < H] \leq \Phi$. We infer from Theorem 2 the existence of little extensions J and L of S with the following properties: $S \leq G \leq J$, $S \leq H \leq L$, Ext $[S < J] =$ Ext $[S < L] = \Phi$. We infer from Theorem 1 the existence of direct decompositions $J = J' \oplus J''$ and $L = L' \oplus L''$ such that J' and L' are equivalent extensions of S. This implies in particular Ext $[S < J'] =$ Ext $[S < L'] = \Phi$. Consequently there exists a group $Z = U \oplus V \oplus W$ where U and J' [and L'] are equivalent extensions of S, $V \cong J''$, $W \cong L''$. It is clear that Z is a little extension of S, since J and L are little extensions of S. Furthermore it is easy to find a subgroup G_0 of $U \oplus V$ such that G and G_0 are equivalent extensions of S; and a subgroup H_0 of $U \oplus W$ such that H and H_0 are equivalent extensions of S. Let $Z_0 = G_0 + H_0$. This is a little extension of S such that Ext $[S < Z_0] \leq \Phi$; and Z_0 is essentially a sum of the extensions G and H of S.

In general it will not be possible just to form $G + H$ with the requirement $S = G \cap H$, since this group need not be a little extension of S.

It should furthermore be noted that the preceding remarks do not settle the [still open] question whether there exists always in $\Lambda(S)$ a least upper bound of any two extension types.

4. Extensions of little groups. So far we have not imposed any restrictions upon the groups S whose extensions were investigated. All the restrictions were imposed upon the extensions themselves. In order that the following extensions of our theory may be attained as simply as possible we restrict now the class of extended groups.

DEFINITION 1. *The group S is a little group, if every Loewy system of S is a little Loewy system of S.*

We mention first that free abelian groups of positive rank, in particular the infinite cyclic groups, are not little groups. If S is such a free abelian group, not 0, then let, for instance, $S = S(0,p)$, $S(3i-2,p) = S(3i-1,p) = S(3i,p) = p^iS$ for $0 < i$ and every prime number p. It is easy to see that these groups $S(j,p)$ form a Loewy system of S, but not a little one; and hence S is not a little group.

A large class of little groups, though not the largest one, has been mentioned in Section 2 [Criterion A]. In particular every finite group is a little group.

Example. Let S be the cyclic group of order p, generated by the element s. Denote by G the group obtained by adjoining to S elements $g(i)$ for $i = 1, 2, \cdots$, subject to the relations $p^i g(i) = s$ for every i. It is easy to see that

$$\text{(L)} \qquad S = S \cap q^j G \text{ for every prime power } q^j.$$

The minimal extension of the little group S which is characterized by the Loewy system (L) is a group of type p^∞ which contains S as its uniquely determined subgroup of order p. But there does not exist a homomorphism of a group of type p^∞ into G which leaves invariant the elements in S. Consequently there does not exist any little extension of the little group S which belongs to the same extension type as G.

Neither subgroups nor quotient groups of little groups need be little. For instance, the additive group of rational numbers is a little group; but its subgroup of integers is not a little group, as has been pointed out before. If furthermore S is the direct sum of a cyclic group Z of order p and of a countable infinity of groups $R(i)$ each of which is isomorphic to the additive group of rational numbers, then S is a little group. If $J(i)$ is the subgroup of the integers in $R(i)$, and if J is the direct sum of the subgroups $J(i)$, then S/J contains an infinity of elements of order p; and now it is easily verified that the quotient group S/J of the little group S is not a little group.

PROPOSITION 1. *If S is a little group, then $\Lambda(S)$ is a complete modular lattice.*

To prove this, it suffices to realize that $\Lambda(S)$ is, by the Characterization Theorem of Section 3, essentially the same as the partially ordered set $L(S)$ of the little Loewy systems of S; and that $L(S)$ is the system of all the Loewy systems of S, since S is a little group. The remainder of the argument is fairly obvious.

If S is a little group, then $\Lambda(S)$ need not satisfy any of the chain conditions. If S happens to be finite, then the descending chain condition holds in $\Lambda(S)$, though the ascending chain condition will, in general, not be satisfied by $\Lambda(S)$.

PROPOSITION 2. *Suppose that the subgroup R of the little group S has the property*

(F) *If S contains only a finite number of elements of order p, then S/R contains only a finite number of elements of order p.*

Then S/R is a little group; and there exists a single valued and order preserving mapping of $\Lambda(S)$ upon $\Lambda(S/R)$ which induces an isomorphism in a suitable subsystem of $\Lambda(S)$.

Proof. Suppose that $T(i,p)$ is a Loewy system of S/R. Then there exists to every prime power p^i a uniquely determined subgroup $S(i,p)$ of S which contains R and satisfies $T(i,p) = S(i,p)/R$. Then $pT(i,p) = [pS(i,p) + R]/R$; and now it is easy to see that the subgroups $S(i,p)$ form a Loewy system of S. Since S is a little group, the $S(i,p)$ form a little Loewy system. Assume now that, for some prime p, the p-Loewy chain $T(i,p)$ is non-trivial. Then the p-Loewy chain $S(i,p)$ cannot be trivial either. Since the $S(i,p)$ form a little Loewy system, this implies that S contains only a finite number of elements of order p; and it follows from Condition (F) that S/R contains only a finite number of elements of order p. Next we infer from the Isomorphism Laws and from Dedekind's Law that

$$\begin{aligned}
&[T(i+2,p) \cap pT(i,p)]/pT(i+1,p) \\
&\quad = ([S(i+2,p)/R] \cap [(pS(i,p)+R)/R])/[(pS(i+1,p)+R)/R] \\
&\quad \simeq [S(i+2,p) \cap (pS(i,p)+R)]/[pS(i+1,p)+R] \\
&\quad = (R + [S(i+2,p) \cap pS(i,p)])/[R + pS(i+1,p)] \\
&\quad \simeq [S(i+2,p) \cap pS(i,p)]/[S(i+2,p) \cap pS(i,p) \cap (R + pS(i+1,p))] \\
&\quad = [S(i+2,p) \cap pS(i,p)]/[pS(i+1,p) + (S(i+2,p) \cap pS(i,p) \cap R)]
\end{aligned}$$

and this last group is clearly a homomorphic image of

$$[S(i+2,p) \cap pS(i,p)]/pS(i+1,p).$$

Since the $S(i,p)$ form a little Loewy system, it follows now that, for every fixed prime p, $[T(i+2,p) \cap pT(i,p)]/pT(i+1,p)$ is finite for every i and 0 for almost every i; and thus we have shown that the $T(i,p)$ form a little Loewy system. Hence S/R is a little group.

It follows from the Characterization Theorem of Section 3 that $\Lambda(S)$ is essentially the same as $L(S)$ and that $\Lambda(S/R)$ is essentially the same as $L(S/R)$. Since S and S/R are little groups, $L(S)$ and $L(S/R)$ are the systems of all Loewy systems in S and S/R respectively.

Denote now by $L_R(S)$ the subset of all the Loewy systems $S(i,p)$ in S which satisfy: $R \leq S(i,p)$ for every prime power p^i.

If $X(i,p)$ is some Loewy system in S, then $X^*(i,p) = [R + X(i,p)]/R$ is a Loewy system in S/R; and it is clear that the mapping of X upon X^* constitutes an order preserving and single valued mapping of $L(S)$ upon $L(S/R)$ which induces an isomorphism of $L_R(S)$ upon $L(S/R)$. This completes the proof.

Appendix I. Generalizations. The possibilities of immediate generalization of the present theory in its entirety seem to be rather meagre. The following extensions of the theory can be effected with hardly any effort; and we enumerate them here because of their possible value in applications.

1. *Extension of the concept of little extension.* If we consider instead of the little extensions defined in Section 1 extensions G of S with the property:

(D) there exists a direct decomposition $G = L \oplus K$ where L is a little extension of S [in the sense of 1, Definition 1];

then it is easy to check that all the results carry over easily from little extensions to extensions with Property (D).

Condition (D) may be applied in the following fashion: Denote by P the subgroup of all the elements in the extension G of S of which a positive multiple belongs to S. If P is a little extension of S, and if G/P is a free abelian group, then P is a direct summand of G and (D) is satisfied. This situation prevails, for instance, in case S and G/S are finitely generated groups.

2. *Subgroups with division.* Suppose that there is given an extension G of the group S, a homomorphism σ of G into the group H and a subgroup K of H such that $K = pK$ for every prime p. Then it is well known that K is a direct summand of H. If $H = K \oplus L$, then $\sigma = \tau + \rho$ where τ is a homomorphism of S into K and ρ is a homomorphism of S into L. It is easy to see that τ is induced by a homomorphism of G into K; and thus σ will be induced by a homomorphism of G into H whenever ρ is induced by a homomorphism of G into L. This simple remark may be used for generalizations of 1, Proposition 1; but because of a certain asymmetry it does not lend itself too well for further use.

3. *Operator groups.* If one imposes sufficiently strong conditions on the operator ring there is no difficulty in extending the whole theory. The following conditions would be obviously sufficient: the ring R of operators is commutative, its ideals are principal, the unique decomposition theorem holds in R, and R is finite modulo its prime ideals.

Appendix II. Dualization. If A is a finite abelian group, then A is isomorphic to its character group; and this isomorphy provides us with a self-duality of A. This self-duality of A interchanges subgroups and quotient groups; and thus it may be used to dualize the results of the present investigation as far as they refer to finite groups only. We shall state the results obtained by this process of dualization without going into the details of their proofs, noting only that the self-duality of A interchanges the subgroup nA and the subgroup $K[n;A]$ of all the solutions of the equation $nx=0$ in A.

I. Suppose that S and T are subgroups of the finite groups A and B respectively.

(a) The homomorphism η of A/S into B/T is induced by a homomorphism of A into B [which maps S into T] if, and only if,

$$[(S+K[p^i;A])/S]\eta \leq (T+K[p^i;B])/T \textit{ for every prime power } p^i.$$

(b) The isomorphism κ of A/S upon B/T is induced by an isomorphism of A upon B [which maps S upon T] if, and only if, $A \simeq B$ and

$$[(S+K[p^i;A])/S]\kappa = (T+K[p^i;B])/T \text{ for every prime power } p^i.$$

If γ is a homomorphism of the group G upon the group $G\gamma = L$, then the pair (G,γ) constitutes *an extension by L.* The extensions (A,α) and (B,β) by the same group L are *equivalent extensions by L,* if there exists an isomorphism σ of A upon B such that $x\alpha = x\sigma\beta$ for every x in A.

II. Suppose that A and B are finite groups. Then the extensions (A,α) and (B,β) by the same group L are equivalent extensions by L if, and only if, $A \simeq B$ and $K[p^i;A]\alpha = K[p^i;B]\beta$ for every prime power p^i.

Extension types by L and their partial ordering may be discussed now in a similar fashion. We omit the details.

UNIVERSITY OF ILLINOIS,
URBANA, ILLINOIS.

BIBLIOGRAPHY.

R. Baer (1), "Splitting endomorphism," *Transactions of the American Mathematical Society*, vol. 61 (1947), pp. 508-516.

A. Kurosch (1), "Kombinatorischer Aufbau der bikompakten topologischen Räume," *Compositio Mathematica*, vol. 2 (1935), pp. 471-476.

(2) "Zur Theorie der teilweise geordneten Systeme von endlichen Mengen," *Recueil Mathématique*, vol. 5 (1939), pp. 343-346.

O. Ore (1), "Structures and group theory. I," *Duke Mathematical Journal*, vol. 3 (1937), pp. 149-174.

N. Steenrod (1), "Universal homology groups," *American Journal of Mathematics*, vol. 58 (1936), pp. 661-701.

GENERALIZED FREE PRODUCTS WITH AMALGAMATED SUBGROUPS.*

By Hanna Neumann.

PART II. The Subgroups of Generalized Free Products.[1]

Introduction. In Part I we defined the generalized free product $\mathfrak{G}$ of the groups $\mathfrak{G}_\alpha$ with amalgamated subgroups $\mathfrak{U}_{\alpha\beta}$. We found that $\mathfrak{G}$ contains a group $\mathfrak{U}$ which is itself a free product with amalgamated subgroups $\mathfrak{U}_{\alpha\beta}$. Its factors are the subgroups $\mathfrak{U}_\alpha$ of $\mathfrak{G}_\alpha$ which are generated by all the amalgamated subgroups $\mathfrak{U}_{\alpha\beta}$ of $\mathfrak{G}_\alpha$. If, in particular, all $\mathfrak{U}_{\alpha\beta}$ are identical, then $\mathfrak{U}_\alpha = \mathfrak{U}_{\alpha\beta}$ for all β, all the groups $\mathfrak{U}_\alpha$ are isomorphic, and $\mathfrak{G}$ is simply the free product of the groups $\mathfrak{G}_\alpha$ with the one amalgamated subgroup $\mathfrak{U}$. In the general case the subgroup $\mathfrak{U}$ of $\mathfrak{G}$ may be loosely described as the smallest generalized free product with amalgamated subgroups $\mathfrak{U}_{\alpha\beta}$.

We are going to show that every subgroup $\mathfrak{H}$ of $\mathfrak{G}$ is itself a generalized free product—unless $\mathfrak{H}$ belongs to $\mathfrak{U}$, in which case nothing can be said.

A. Kurosch [1] [2] has shown that every subgroup $\mathfrak{H}(\mathfrak{F})$ of a free product $\mathfrak{F}$ is itself a free product. The proof consists in the construction of a suitable system of generators for $\mathfrak{H}(\mathfrak{F})$. In Part I (1) we obtained the generalized free product $\mathfrak{G}$ of the groups $\mathfrak{G}_\alpha$ as a homomorphic image of the free product $\mathfrak{F}$ of these groups $\mathfrak{G}_\alpha$. Every subgroup $\mathfrak{H}$ of $\mathfrak{G}$ will, therefore, be the map of a certain subgroup $\mathfrak{H}(\mathfrak{F})$ of $\mathfrak{F}$. Using Kurosch's result, it is not difficult to see that the homomorphism which maps $\mathfrak{F}$ onto $\mathfrak{G}$, will map the free factors of $\mathfrak{H}(\mathfrak{F})$ onto groups generating $\mathfrak{H}$ in such a way that certain subgroups of the free factors of $\mathfrak{H}(\mathfrak{F})$ are amalgamated. The difficulty is to show that $\mathfrak{H}$ is actually the *free* product of these factors with amalgamated subgroups, not only a homomorphic image of it. This difficulty necessitated a closer examination of the generators of $\mathfrak{H}(\mathfrak{F})$ as used by Kurosch and the relations that may arise between their maps. Thus I was led to the construction of a system of generators for the subgroup $\mathfrak{H}$ of $\mathfrak{G}$, in principle similar to Kurosch's generators, but selected with greater care to meet our particular needs.

---

* Received August 27, 1947.

[1] Part I. Definitions and general properties. This Journal, vol. 70 (1948), pp. 590-625. Results will be quoted thus: I. 5. 4, meaning Theorem 5. 4 of Part I.

[2] Numbers in square brackets refer to the list of references at the end of this paper.

Throughout the paper, we assume $\mathfrak{G}$ to be a generalized free product. The (slight) simplifications in the case of a free product with one amalgamated subgroup are indicated briefly in square brackets.

The main tool is the length of an element of $\mathfrak{G}$ as defined in I. 5. 1., in particular, comparison of the length of a product with that of its factors. We begin, therefore, with a restatement of the fundamental properties concerning the representation and multiplication of elements of $\mathfrak{G}$ (**1**), followed by a number of useful facts and definitions in connection with the length of a product in terms of that of its factors (**2**).

In **3** we define a system of generators for a given subgroup $\mathfrak{H}$ of $\mathfrak{G}$. The aim of the subsequent paragraphs is to obtain a comparison of the length of an element of $\mathfrak{H}$ with the length of the generators involved in its representation. This leads to necessary conditions for a word in our generators to be shorter than the longest generator which occurs in the word (**7**), and hence to a description of all possible relations between these generators (**8**). These relations, however, are not yet sufficiently simple. But the properties of the generators derived so far show the way to a further improvement in their original choice, with the result that the relations are considerably simplified (**11, 12**), and can now be interpreted so as to give the main result (**13**):

$\mathfrak{H}$ is a generalized free product with amalgamated subgroups. The factors are either subgroups of conjugates of the groups $\mathfrak{G}_\alpha$, or of $\mathfrak{U}$ (the latter become trivial in the case of a free product with one amalgamated subgroup); or they are of a type analogous to the free generators which occur in Kurosch's representation of the subgroup of a free product. The amalgamated subgroups are generated by subgroups of conjugates of the groups $\mathfrak{U}_\alpha$.

If $\mathfrak{G}$ is the free product of the groups $\mathfrak{G}_\alpha$ with one amalgamated subgroup $\mathfrak{U}$, and if $\mathfrak{U}$ is self-conjugate in $\mathfrak{G}$, (i. e. in every factor $\mathfrak{G}_\alpha$), then $\mathfrak{H}$ is itself a free product with one amalgamated subgroup $\mathfrak{V}$. $\mathfrak{V}$ is simply the meet of $\mathfrak{H}$ and $\mathfrak{U}$. This case includes in particular results by Kurosch and Kalaschnikov (cf. [2]), where $\mathfrak{U}$ is assumed to be a subgroup of the centre of every factor $\mathfrak{G}_\alpha$.

In **14** we investigate to what extent those factors of $\mathfrak{H}$ which are conjugates of subgroups of $\mathfrak{G}_\alpha$, or of subgroups of $\mathfrak{U}$, are uniquely determined by $\mathfrak{H}$. The results are again a direct generalization of the corresponding results obtained by Kurosch [1].

We conclude with some remarks on the special part played by the subgroup $\mathfrak{U}$ of $\mathfrak{G}$ in the general case, indicating the reason why similar results cannot be expected for the subgroups of this "smallest" generalized free product $\mathfrak{U}$.

1. Let $\mathfrak{G} = \{\overset{*}{\prod_\alpha} \mathfrak{G}_\alpha;\ \mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}\}$ be the generalized free product of the groups $\mathfrak{G}_\alpha$ (finite or infinite in number) with amalgamated subgroups $\mathfrak{U}_{\alpha\beta}$. I. e. $\mathfrak{G}$ is generated by the groups $\mathfrak{G}_\alpha$, and any two of them, $\mathfrak{G}_\alpha$ and $\mathfrak{G}_\beta$, have meet $\mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}$ in $\mathfrak{G}$. As before, we denote by $\mathfrak{U}_\alpha$ the group generated by all the subgroups $\mathfrak{U}_{\alpha\beta}$ (α fixed) of $\mathfrak{G}_\alpha$, and by $\mathfrak{U}$ the subgroup of $\mathfrak{G}$ which is generated by all the groups $\mathfrak{U}_\alpha$. Then $\mathfrak{U}$ is the generalized free product of the groups $\mathfrak{U}_\alpha$,

$$\mathfrak{U} = \{\overset{*}{\prod_\alpha} \mathfrak{U}_\alpha;\ \mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}\}.$$

and the meet of $\mathfrak{U}$ and $\mathfrak{G}_\alpha$ is $\mathfrak{U}_\alpha$, for every α. In fact, it follows from I. 7. 0 and I. 6. 02 that $\mathfrak{U}$ and $\mathfrak{G}_\alpha$ generate in $\mathfrak{G}$ a group which is the free product of $\mathfrak{U}$ and $\mathfrak{G}_\alpha$ with the amalgamated subgroup $\mathfrak{U}_\alpha$.

We assume throughout the paper that for at least one suffix α, $\mathfrak{G}_\alpha$ is actually larger than $\mathfrak{U}_\alpha$; hence $\mathfrak{G}$ is actually larger than $\mathfrak{U}$. This restriction means that no results will be obtained on the subgroups of the smallest generalized free product $\mathfrak{U}$. We touch upon this fact once more in the concluding paragraph.

Every element G of $\mathfrak{G}$ is of the form

$$G = U_0 G_{\alpha_1} U_1 G_{\alpha_2} \cdots G_{\alpha_l} U_l, \qquad l \geq 0, \tag{1.0}$$

with U_λ in $\mathfrak{U}$ ($\lambda = 0, 1, \cdots, l$), G_{α_λ} in $\mathfrak{G}_{\alpha_\lambda} - \mathfrak{U}_{\alpha_\lambda}$ ($\lambda = 1, \cdots, l$), and if $\alpha_\lambda = \alpha_{\lambda+1}$, then U_λ in $\mathfrak{U} - \mathfrak{U}_{\alpha_\lambda}$ ($\lambda = 1, \cdots, l$). This representation is not unique: The "factors" G_{α_λ} are determined but for left- and right-hand factors in $\mathfrak{U}_{\alpha_\lambda}$, and the "$\mathfrak{U}$-factors" U_λ are determined but for left-hand factors in $\mathfrak{U}_{\alpha_\lambda}$ (for $\lambda = 1, \cdots, l$) and right-hand factors in $\mathfrak{U}_{\alpha_{\lambda+1}}$ (for $\lambda = 0, \cdots, l-1$). But the number $l = l(G)$ of factors is an invariant of G. We call it again the length of G. Obviously, the formal inverse of the representation 1. 0 is itself of the form 1. 0; hence we have $l(G^{-1}) = l(G)$.

The elements of length zero are exactly the elements of $\mathfrak{U}$. But it should be noted that the elements of length one do not necessarily belong to a factor $\mathfrak{G}_\alpha$; they do if, and only if, both U_0 and U_1 belong to $\mathfrak{U}_\alpha$.

[If $\mathfrak{G}$ is the free product with one amalgamated subgroup, $\mathfrak{U}$ belongs to every factor $\mathfrak{G}_\alpha$, and 1. 0 simplifies to $G = U G_{\alpha_1} G_{\alpha_2} \cdots G_{\alpha_l}$ ($\alpha_\lambda \neq \alpha_{\lambda+1}$). Here G belongs to a factor $\mathfrak{G}_\alpha$ if, and only if, $l(G) \leq 1$.]

The representation 1. 0 can be made unique, e. g. if we use the normal form I. 5. 1. But the use of some such normal form will, in general, not be very helpful in this part. What matters is the way in which two such elements are multiplied, in particular the length of the product in relation to that of the factors.

Let

$$G = U_0 G_{\alpha_1} U_1 \cdots G_{\alpha_l} U_l$$

and

$$H = V_0 H_{\beta_1} V_1 \cdots H_{\beta_m} V_m$$

be two elements of length l and m respectively. If one of them, G say, is of length zero, we have $l(GH) = l(HG) = l(H)$. If both elements are of length ≥ 1, we have

$$GH = U_0 G_{\alpha_1} U_1 \cdots G_{\alpha_l} (U_l V_0) H_{\beta_1} V_1 \cdots H_{\beta_m} V_m.$$

If either $\alpha_l \neq \beta_1$, or $\alpha_l = \beta_1$, but in that case $U_l V_0$ does not belong to $\mathfrak{U}_{\alpha_l} = \mathfrak{U}_{\beta_1}$, then this is a representation of the form 1. 0 for the product GH; hence $l(GH) = l(G) + l(H)$. If, however, $\alpha_l = \beta_1 = \alpha$, and $U_l V_0 = U_\alpha$ belongs to $\mathfrak{U}_\alpha$, then $G_{\alpha_l}(U_l V_0) H_{\beta_1}$ is an element of $\mathfrak{G}_\alpha$. Either it belongs to $\mathfrak{G}_\alpha - \mathfrak{U}_\alpha$; in that case we say that $G_{\alpha_l} U_l$ and $V_0 H_{\beta_1}$ 'amalgamate modulo $\mathfrak{U}$,' or briefly, that they amalgamate. Or, the product $G_{\alpha_l}(U_l V_0) H_{\beta_1}$ belongs to $\mathfrak{U}$, i. e. to $\mathfrak{U}_\alpha$. Then we say, $G_{\alpha_l} U_l$ and $V_0 H_{\beta_1}$ 'cancel modulo $\mathfrak{U}$,' or simply, they cancel. In that case, further amalgamation or cancellation may take place, if also $G_{\alpha_{l-1}}$ and H_{β_2} belong to the same group, $\mathfrak{G}_\beta$ say, and the product of the $\mathfrak{U}$-factors between them belongs to the corresponding subgroup $\mathfrak{U}_\beta$ of $\mathfrak{U}$. Otherwise, no further amalgamation, or cancellation, is possible.

Thus, a representation 1. 0 of the product GH is obtained after a number r $(0 \leq r \leq \min(l, m))$ of cancellations followed, possibly, by one amalgamation. Every cancellation disposes of two of the $l + m$ factors of G and H, the amalgamation replaces two factors by one. Hence:

1. 10 $$l(GH) = l(G) + l(H) - 2r - \epsilon,$$

where $\epsilon = 0$ if no amalgamation takes place, $\epsilon = 1$ otherwise. In either case, if in the product GH there are exactly s factors at the beginning of G which are not touched by cancellation and amalgamation, then $l(G) = s + r + \epsilon$, hence

1. 11 $$l(GH) = l(H) + s - r,$$

and therefore

1. 12 $$l(GH) \gtreqless l(H) \text{ according as } s \gtreqless r.$$

An analogous formula compares the length of the product with that of the first factor G.

That in the product GH r factors cancel, means that

$$\alpha_l = \beta_1, \alpha_{l-1} = \beta_2, \cdots, \alpha_{l-r+1} = \beta_r,$$

and for every ρ $(1 \leq \rho \leq r)$ we have

1.21 $(G_{\alpha_{l-\rho+1}} U_{l-\rho+1} \cdots G_{\alpha_l} U_l)(V_0 H_{\beta_1} \cdots V_{\rho-1} H_{\beta_\rho}) = U_{\beta_\rho}$ $(U_{\beta_\rho}$ in $\mathfrak{U}_{\beta_\rho})$;

and if one further factor amalgamates, then also $\alpha_{l-r} = \beta_{r+1}$, and

1.22 $(G_{\alpha_{l-r}} U_{l-r} \cdots G_{\alpha_l} U_l)(V_0 H_{\beta_1} \cdots V_r H_{\beta_{r+1}}) = F_{\beta_{r+1}}(F_{\beta_{r+1}}$ in $\mathfrak{G}_{\beta_{r+1}} - \mathfrak{U}_{\beta_{r+1}})$.

[In the case of a free product with one amalgamated subgroup, cancellation or amalgamation will take place in the product GH if, and only if, $\alpha_l = \beta_1$. The essential formulae 1.10, 1.11, and 1.12 remain, of course, the same.]

2. Let G and H be two elements of length l and m respectively, where $l \leq m$. Throughout the following paragraphs we shall be particularly interested in the case that

2.01 $$l(G^{-1}H) \leq l(H),$$

or

2.02 $$l(HG^{-1}) \leq l(H).$$

Both inequalities are always true, if $l(G) = 0$; in that case, of course, $l(G^{-1}H) = l(HG^{-1}) = l(H)$. If $l = l(G) \geq 1$, they imply, by 1.12, that at least as many factors of G^{-1} cancel as remain untouched, i. e. at least $[\frac{1}{2}l]$ factors cancel, and, if l is odd, at least one more factor cancels or amalgamates. If, in 2.01 or 2.02, we are to have equality, then exactly $[\frac{1}{2}l]$ factors of G^{-1} must cancel, and in the case of even l nothing amalgamates, while in the case of odd l one more factor of G^{-1} amalgamates, but does not cancel.

Now the formal inverse of a representation 1.0 of G gives us a representation 1.0 of G^{-1}. Hence, if again

$$G = U_0 G_{\alpha_1} U_1 \cdots G_{\alpha_l} U_l \text{ and } H = V_0 H_{\beta_1} V_1 \cdots H_{\beta_m} V_m,$$

and if in the product $G^{-1}H$ r factors cancel, we have by 1.21,

$$(G_{\alpha_r}^{-1} \cdots G_{\alpha_1}^{-1} U_0^{-1})(V_0 H_{\beta_1} \cdots H_{\beta_r}) = U_{\beta_r}(U_{\beta_r} \text{ in } \mathfrak{U}_{\beta_r});$$

and therefore, as $\beta_r = \alpha_r$, the representation

$$H = U_0 G_{\alpha_1} \cdots G_{\alpha_r}(U_{\beta_r} V_r) H_{\beta_{r+1}} \cdots H_{\beta_m} V_m$$

is also of the form 1.0; i. e. representations 1.0 of G and H may be assumed to be identical up to the r-th factor inclusive. If in the product $G^{-1}H$ one further factor amalgamates, one concludes similarly from 1.22, that the first r factors of G and H, including the following $\mathfrak{U}$-factor, may be assumed to be identical, and the $(r+1)$-st factors of G and H belong to the same group $\mathfrak{G}_{a_{r+1}}$.

If, in particular, G and H satisfy 2.01, then $r \geq [\frac{1}{2}l]$, i. e. at least the first $[\frac{1}{2}l]$ factors of G may be taken as identical to as many factors in the beginning of H, and, if l is odd, one more factor of G certainly belongs to the same group as the corresponding factor of H. And 2.02 entails a similar statement for the last factors of G and H respectively. This leads to the following way of writing a representation 1.0 of an element G of length l:

If $l = 2k + 1$ is odd, and

$$G = U_0G_1 \cdots G_kU_kG_{k+1}U_{k+1} \cdots G_{2k+1}U_{2k+1},$$

we call $S = S(G) = U_0G_1 \cdots G_kU_k$ the first half of G,

$T = T(G) = U_{k+1}G_{k+2} \cdots G_lU_l$ the second half of G,

$C = C(G) = G_{k+1}$ the central factor of G;

if $l = 2k$ is even, and

$$G = U_0G_1 \cdots U_{k-1}G_kU_kG_{k+1} \cdots G_lU_l,$$

we call $S = S(G) = U_0G_1 \cdots U_{k-1}G_k$ the first half of G,

$T = T(G) = G_{k+1} \cdots G_lU_l$ the second half of G,

$C = C(G) = U_k$ the central factor of G.

Then, in both cases, the representation 1.0 of G and G^{-1} respectively is of the form $G \equiv SCT$ and $G^{-1} \equiv T^{-1}C^{-1}S^{-1}$,[3] where $S = S(G)$ and $T = T(G)$ are elements of length $k = [\frac{1}{2}l]$, and $C = C(G)$ of length one or zero according as l is odd or even. [It should always be remembered, that if l is odd, $S = S(G)$ includes the $\mathfrak{U}$-factor which follows the factor G_k, while in the case of even l, S is the partial product of G including G_k, but not U_k. And correspondingly for T. No such distinction occurs in the case of a free product with one amalgamated subgroup.]

With this notation, $l(G^{-1}H) \leq l(H)$ means that G can be written in such a way in the form $G = SCT$, that at least S is identical with as many factors in the beginning of H, and if l is odd, C belongs to the same group, as the next following factor of H. Similarly for T, if $l(HG^{-1}) \leq l(H)$. In particular,

---

[3] Cf. Kurosch, [1], p. 650.

2.03. *If* $l(G) = l(H) = l$, *and* $l(G^{-1}H) \leq l$, *then* $G = S(G)C(G)T(G)$, $H = S(H)C(H)T(H)$, *where we may assume* $S(G) \equiv S(H)$; and $C(G)$ *and* $C(H)$ *belong to the same group. Correspondingly, if* $l(HG^{-1}) \leq l$, *it may be assumed that* $T(G) \equiv T(H)$, *and* $C(G)$ *and* $C(H)$ *belong to the same group.*

If, under the same assumptions, $l(G^{-1}H) = l$, then if l is odd, $C(G)^{-1}$ and $C(H)$ amalgamate, but do not cancel; if l is even, $C(G)^{-1}$ and $C(H)$ do, of course, amalgamate, but the first factors of $T(G)$ and $T(H)$ are not even amalgamated in the product $G^{-1}H = T(G)^{-1}C(G)^{-1}C(H)T(H)$. Similarly, if $l(HG^{-1}) = l$.

The following lemma, frequently used later on, is an immediate consequence of the preceding discussion:

2.1. *Let* F, G, H *be three elements of length* k, l, m *respectively, and* $k \leq l \leq m$. *Then*

2.11. $l(F^{-1}G) \leq l(G)$ *and* $l(G^{-1}H) \leq l(H)$ *implies that also* $l(F^{-1}H) \leq l(H)$;

2.12. $l(F^{-1}H) \leq l(H)$ *and* $l(G^{-1}H) \leq l(H)$ *implies that also* $l(F^{-1}G) \leq l(G)$.

Also:

2.2. *Let again* F, G, H *be three elements of length* k, l, m *respectively, and* $k \leq l \leq m$. *Then* $l(GF^{-1}) = l(G)$ *and* $l(GF^{-1}H) \leq l(H)$ *implies* $l(F^{-1}H) \leq l(H)$.

Proof. 2.2 follows from 2.11 in the following way: Put $FG^{-1} = G'$; then $l(G'^{-1}) = l(G') = l(G) = l$. Moreover, we have $l(F^{-1}G') = l(G^{-1}) = l(G')$ and $l(G'^{-1}H) = l(GF^{-1}H) \leq l(H)$; hence, 2.11 applied to the elements F, G', H gives 2.2.

To complete the preliminaries, we define:

2.30. *An element* K *such that* $l(K^2) \leq l(K)$, *is called a transform.*

We apply 2.03, with $G = K^{-1}$, and $H = K$; then $l(G^{-1}H) \leq l(H)$ and $l(HG^{-1}) \leq l(H)$, hence K can be written in the form $K \equiv S(K)C(K)T(K)$ in such a way that $T(K) \equiv S(K)^{-1}$, i.e.

2.31. $K = S(K)C(K)S(K)^{-1}$ *is a transform of its central factor; i.e. a transform of an element of a group* $\mathfrak{G}_a$, *if* $l(K)$ *is odd, and a transform of an element of* $\mathfrak{U}$, *if* $l(K)$ *is even.*

2. 32. *If K' is another transform of the same length as K, and if $l(K'K) \leq l(K)$, then $K' = S(K)C'S(K)^{-1}$, where $C' = C(K')$ belongs to the same group as $C = C(K)$; and we have also $l(KK') \leq l(K)$.*

This follows immediately from 2. 03.

If, in 2. 32, $l(KK') < l(K)$, then, in case $l(K)$ is odd, the central factors of K and K' must cancel. As they both belong to $\mathfrak{G}_a$, we have $CC' = U_a$, hence $KK' = S(K)U_aS(K)^{-1}$. This is again a transform, but of length $\leq l(K) - 1$. The exact length depends on whether or not further factors of $S(K)$ and $S(K)^{-1}$ cancel or amalgamate.

If $l(K)$ is even, we have $K = S(K)US(K)^{-1}$, $K' = S(K)U'S(K)^{-1}$ and $l(KK') < l(K)$ implies that in the product $KK' = S(K)UU' \cdot S(K)^{-1}$ at least one cancellation or amalgamation takes place between $S(K)$ and $S(K)^{-1}$. If the last factor of $S(K)$ belongs to the group $\mathfrak{G}_\beta$, UU' must, therefore, be an element U_β of $\mathfrak{U}_\beta$, and $KK' = S(K)U_\beta S(K)^{-1}$.

We define:

An element of $\mathfrak{G}$ which is conjugate to an element of some group $\mathfrak{U}_a$, is called a $\mathfrak{U}$-transform.

Then we have seen:

2. 33. *If $K = SCS^{-1}$ and $K' = SC'S^{-1}$ are two transforms of the same length l such that $l(KK') < l$, then $KK' = SU_\gamma S^{-1}$ is a $\mathfrak{U}$-transform. The suffix γ is that of the group of the central factors of K and K', if l is odd, and that of the group of the last factor of S, if l is even.*

Finally, we add a useful criterion for an element to be a transform:

2. 4. *If G is any element of length l, K an element of length $\leq l$, then $l(KG) \leq l$ and $l(K^{-1}G) \leq l$ can hold simultaneously if, and only if, K is a transform.*

Proof. If K is a transform, then with $l(KG) \leq l$, certainly also $l(K^{-1}G) \leq l$, by 2. 03 and 2. 31. If on the other hand, both inequalities are satisfied, let $K = S(K)C(K)T(K)$, $l(K) = l_1$. As $l(KG) \leq l$, $T(K)^{-1}$ may be assumed to be identical with the first $[\frac{1}{2}l_1]$ factors of $S(G)$, and, if l_1 is odd, $C(K)$ will belong to the same group as the following factor of G. As also $l(K^{-1}G) \leq l$, $S(K)$ may be taken as identical with the first $[\frac{1}{2}l_1]$ factors of $S(G)$; i. e. $S(K)$ and $T(K)^{-1}$ may be assumed to be identical in a representation 1. 0 of K. Hence K is a transform.

[A transform of length one need not be an element of a group $\mathfrak{G}_a$; it may be of the form UG_aU^{-1} (U not in $\mathfrak{U}_a$). But in the case of a free product with one amalgamated subgroup, a transform of length one is necessarily an element of a group $\mathfrak{G}_\alpha$. Moreover, a transform can only be of odd length, i. e. it is always the transform of an element of a group $\mathfrak{G}_\alpha$, not of $\mathfrak{U}$.]

3. Let $\mathfrak{H}$ be any subgroup of $\mathfrak{G}$ which is not wholly contained in $\mathfrak{U}$; its meet with $\mathfrak{U}$ we denote by $\mathfrak{V}$.

We choose a system of generators for $\mathfrak{H}$ as follows:[4] by ϕ_0 we denote the set of all elements V of $\mathfrak{V}$. Every element of ϕ_0 is taken as a generator.

If, for all ordinal numbers $\sigma' < \sigma$, the sets $\phi_{\sigma'}$ of generators are chosen, let $\mathfrak{K}_\sigma$ be the subgroup of $\mathfrak{H}$ which is generated by all the generators of all the sets $\phi_{\sigma'}$ $(\sigma' < \sigma)$ taken together. If $\mathfrak{K}_\sigma$ is not yet identical with $\mathfrak{H}$, we define the set ϕ_σ of generators as follows:

Let l be the length of a shortest element in $\mathfrak{H} - \mathfrak{K}_\sigma$. If amongst the elements of length l of $\mathfrak{H} - \mathfrak{K}_\sigma$ there exist some which are transforms, we choose one of those as first generators t_σ of the set ϕ_σ; and in addition, we take into ϕ_σ all other transforms t'_σ of $\mathfrak{H} - \mathfrak{K}_\sigma$ which are of the same length l and such that $l(t'_\sigma t_\sigma) \leq l$. By 2. 32, all these transforms may be written with identical first and second halves, i. e. we start the set ϕ_σ with all elements of length l of $\mathfrak{H} - \mathfrak{K}_\sigma$ which are of the form

3. 01. $t_\sigma = S_\sigma C_\sigma C_\sigma^{-1}$, $t'_\sigma = S_\sigma C'_\sigma S_\sigma^{-1}, \cdots$, *where all central factors belong to the same group.*

If outside the group generated by $\mathfrak{K}_\sigma$ and all the elements $t_\sigma, t'_\sigma, \cdots$ there are elements f of $\mathfrak{H}$ of the same length l and with the property $l(f^{-1}t_\sigma) \leq l$, we choose such an element as next generator $f_{\sigma,1}$ of ϕ_σ. I. e. $f_{\sigma,1}$ is determined by

3. 02 $$l(f_{\sigma,1}) = l \text{ and } l(f_{\sigma,1}{}^{-1}t_\sigma) \leq l.$$

If outside the group generated by $\mathfrak{K}_\sigma$ and the generators $t_\sigma, t'_\sigma, \cdots$ and $f_{\sigma,1}$ there is still an element f in $\mathfrak{H}$ which is of length l and such that $l(f^{-1}t_\sigma) \leq l$, then we choose this as next generator $f_{\sigma,2}$ of ϕ_σ, and so we go on. This process will reach its end at a certain ordinal number τ_σ, when outside the group generated by $\mathfrak{K}_\sigma$, $t_\sigma, t'_\sigma, \cdots, f_{\sigma\tau}$ $(1 \leq \tau < \tau_\sigma)$ no more elements f of $\mathfrak{H}$ exist which are of length l and such that $l(f^{-1}t_\sigma) \leq l$. Then we denote by ϕ_σ the set of generators $t_\sigma, t'_\sigma, \cdots$, and $f_{\sigma,\tau}$ $(1 \leq \tau < \tau_\sigma)$ and by $\mathfrak{K}_{\sigma+1}$ the subgroup of $\mathfrak{H}$ which is generated by $\mathfrak{K}_\sigma$ and ϕ_σ.

If, on the other hand, there is no transform of length l in $\mathfrak{H} - \mathfrak{K}_\sigma$, we choose any element of length l as first generator $f_{\sigma,1}$ of ϕ_σ, and then go on as before; ϕ_σ will then consist of generators $f_{\sigma,\tau}$ $(1 \leq \tau < \tau_\sigma)$ only, and, for every τ,

---

[4] Cf. Kurosch, [1], pp. 651-652.

$$l(f_{\sigma\tau}) = l \text{ and } l(f_{\sigma\tau}^{-1}f_{\sigma 1}) \leq l. \tag{3.03}$$

$\mathfrak{K}_{\sigma+1}$ is defined as before, as the group generated by $\mathfrak{K}_\sigma$ and ϕ_σ.

This process of generating $\mathfrak{H}$ will reach its end at a certain ordinal number σ_0 when $\mathfrak{K}_{\sigma_0} = \mathfrak{H}$. Then all the elements of all the sets ϕ_σ $(0 \leq \sigma < \sigma_0)$ form the system ϕ of generators of $\mathfrak{H}$.

From the discussions in **2** it is seen that the generators of each set ϕ_σ are all of the same length, have the same first halves, and, if of odd length, their central factors at least amalgamate; i. e. all the generators of ϕ_σ can be written as:

$$t_\sigma = S_\sigma C_\sigma S_\sigma^{-1},\ t'_\sigma = S_\sigma C'_\sigma S_\sigma^{-1}, \cdots, f_{\sigma\tau} = S_\sigma C_{\sigma\tau} T_{\sigma\tau}, \tag{3.10}$$

where all the central factors $C_\sigma, C'_\sigma, \cdots,$ and $C_{\sigma\tau}$ belong to $\mathfrak{U}$, if l is even, or to $\mathfrak{H}_a - \mathfrak{U}_a$, if l is odd. In particular we have

3. 11. *$l(g_i^{-1}g_k) \leq l$ for any two generators g_i and g_k of ϕ_σ.*

Generators of the same length need not necessarily belong to the same set ϕ_σ; but:

3. 12. *If f is an element of the same length l as the generators of ϕ_σ, such that for some generator g of ϕ_σ $l(f^{-1}g) \leq l$, then f belongs to $\mathfrak{K}_{\sigma+1}$. If, in addition, f is a transform, then f is itself a generator of the form t_σ.*

Proof. As $l(f^{-1}g) \leq l$, it follows from 2. 03 that f is of the form $f = S_\sigma CT$ where C belongs to the same group as the central factors of the generators of ϕ_σ. Hence, for every generator $g' = t_\sigma$ or $= f_{\sigma\tau}$ of ϕ_σ we have $l(f^{-1}g') \leq l$. By definition of ϕ_σ, no elements of length l with this property can be found in $\mathfrak{H} - \mathfrak{K}_{\sigma+1}$. Hence f belongs to $\mathfrak{K}_{\sigma+1}$.

If f is a transform, then f is of the form $f = S_\sigma C S_\sigma^{-1}$, and as it is of length l, and all such elements were taken into ϕ_σ, $f = t_\sigma$.

The process of choosing the generators defines an order relation between them. If, in general, we denote unspecified generators, and their inverses, by $g, g_i, g_k, \cdots$, then g_i is earlier than g_k if in the process of choosing the generators g_i or its inverse was chosen before g_k or its inverse. Unless both, g_i and g_k, are of the form $g_i = t_\sigma$, $g_k = t'_\sigma$, in which case no order relation is given for them, g_i is either earlier or later than g_k whenever $g_i \neq g_k^\epsilon$ $(\epsilon = \pm 1)$:

g_i is earlier than g_k, if $g_i^{\epsilon_i}$ belongs to ϕ_{σ_i}, and $g_k^{\epsilon_k}$ belongs to ϕ_{σ_k}, where $\sigma_i < \sigma_k$. If $\sigma_i = \sigma_k$, i. e. g_i and g_k belong to the same set ϕ_σ, then g_i is earlier than g_k if either $g_i = t_\sigma$, $g_k^\epsilon = f_{\sigma\tau}$, or if $g_i^{\epsilon_i} = f_{\sigma\tau_i}$ and $g_k^{\epsilon_k} = f_{\sigma\tau_k}$ where $\tau_i < \tau_k$. If, in particular, g_i and g_k are of different length, then the shorter generator is earlier than the longer one.

It will be noted that the rules governing the choice of this system ϕ of generators for $\mathfrak{H}$ leave us, in general, considerable freedom in every step. This freedom of choice will be used later to impose a further condition on the generators in order to simplify the possible relations between them. But we have to learn more about the properties of these generators, before we can profitably choose between different sets obtainable by the process described above.

4. This and the following paragraph deal with the immediate consequences of this choice of generators. They are very much the same as in the simpler case of a free product of groups where Kurosch [1] and Neumann [3] use similar systems of generators.

Our principal aim is to derive a system of defining relations for these generators. As usual, we call $P = \prod_1^r g_\rho$ a word in the generators $g_1, \cdots, g_r$, if no trivial cancellations are possible, i. e. if $g_\rho \neq g_{\rho+1}^{-1}$ for $\rho = 1, \cdots, r-1$. A relation is a word R which is the unit element as element of $\mathfrak{H}$, i. e. it is of length zero. In order to find all possible relations, we shall investigate under what conditions a word P in generators g_ρ of length $l(g_\rho) \leq l$ can be shorter than l, i. e. shorter than the longest generator which occurs in the word.

In this paragraph we deal with words in generators which are all of equal length l. We shall need the almost obvious lemma:

4. 0. *If h is an element of length l in $\mathfrak{H}$, then h can be expressed by means of generators of length $\leq l$ only.*

Proof. Let σ be the first ordinal number so that ϕ_σ consists of generators of length $l_1 > l$. $\mathfrak{K}_\sigma$ is, by definition, the subgroup of $\mathfrak{H}$ which is generated by all generators earlier than the first generator of ϕ_σ, i. e. by all generators of length $\leq l$, and the minimum length of the elements of $\mathfrak{H} - \mathfrak{K}_\sigma$ is l_1. As h is of length l, h belongs to $\mathfrak{K}_\sigma$.

4. 1. *If g_i and g_k are generators of equal length l, and if g_i is earlier than g_k, then $l(g_ig_k) \geq l$ and $l(g_kg_i) \geq l$.*

Proof. If e. g. $l(g_ig_k) < l$, and if g_i belongs to ϕ_σ, then, by 4. 0, the product $g_i g_k$ certainly belongs to $\mathfrak{K}_\sigma$. Hence g_k belongs to the group generated by $\mathfrak{K}_\sigma$ and all generators of ϕ_σ which are not later than g_i, contrary to the assumption that g_k is later than g.

A product of two generators of the same length l can, therefore, be shorter than l only, if no order relation is defined for g_i and g_k, i. e. if they belong to the same set ϕ_σ and, unless $g_k = g_i^{-1}$, both are of the form $g_i = t_\sigma$, $g_k = t'_\sigma$. And then $l(g_i g_k) < l$ can, in fact, happen; by 2.33, their product is then a $\mathfrak{U}$-transform, $g_i g_k = S_\sigma U_a S_\sigma^{-1}$.

In all other cases, i. e. when one of the generators g_i and g_k is earlier than the other, we need to know when their product can be of minimal length $l(g_i g_k) = l(g_i) = l(g_k) = l$. This certainly is the case, if g_i and g_k belong to the same set ϕ_σ, and either $g_i = t_\sigma$, $g_k = f_{\sigma\tau}$, or $g_i = f_{\sigma\tau_i}{}^{-1}$, $g_k = f_{\sigma\tau_k}$. In both cases, by 3.11, $l(g_i g_k) \leq l$. But, by 4.1, the product is at least of length l, so that it follows now, that for any two generators $g_i = t_\sigma$, $g_k = f_{\sigma,\tau}$ or $g_i = f_{\sigma\tau_i}{}^{-1}$, $g_k = f_{\sigma\tau_k}$ $(\tau_i \neq \tau_k)$ we have $l(g_i g_k) = l$. We want to show that these are in fact the only cases; in all other cases we have $l(g_i g_k) > l$.

In the next two lemmas, g_i and g_k denote generators of certain sets ϕ_{σ_i} and ϕ_{σ_k}, but not their inverses, so that $g_i = t_{\sigma_i}$ or $g_i = f_{\sigma_i,\tau}$, and correspondingly for g_k.

4.2.[5] *If $l(g_i) = l(g_k) = l$, and $l(g_i g_k^{-1}) = l$, then also $l(g_i^{-1} g_k) = l$ and $g_i = t_\sigma$, $g_k = t'_\sigma$ are two transforms belonging to the same set ϕ_σ.*

COROLLARY. *If $g_i = f_{\sigma\tau}$, and g is any other generator of the same length, then $l(g_i g) > l(g_i)$ and $l(g_i g^{-1}) > l(g_i)$.*

Proof. Assume that one of the generators, g_i say, is earlier than the other.

If g_i and g_k belong to different sets, g_i to ϕ_{σ_i}, g_k to ϕ_{σ_k}, then $\sigma_i < \sigma_k$. We put $f = g_i g_k^{-1}$. Then $l(f) = l$, and $f^{-1} g_i = g_k$ is of length l. Hence, by 3.12, $f = g_i g_k^{-1}$ belongs to $\mathfrak{K}_{\sigma_i+1}$. As also g_i belongs to the group $\mathfrak{K}_{\sigma_i+1}$, it follows that also g_k belongs to it, contrary to the assumption that g_k belongs to a later set than g_i.

Hence, g_i and g_k belong to the same set ϕ_σ. As g_k is later than g_i in ϕ_σ, g_k is of the form $f_{\sigma\tau_k}$, and $g_i = t_\sigma$ or $= f_{\sigma\tau_i}$. Also $f = g_i g_k^{-1}$ is again of length l (otherwise g_k were not later than g_i), and $l(f g_k) = l$; therefore, again by 3.12, $f = t'_\sigma$ is itself a generator of ϕ_σ which is earlier than g_k. But $g_k = f^{-1} g_i$, contrary to the assumption that g_k is a generator later than g_i.

Hence, neither of the two is earlier than the other, i. e. $g_i = t_\sigma$, $g_k = t'_\sigma$.

---

[5] Cf. Kurosch [1], p. 653; Neumann [3], p. 14.

4. 3.[5] *If* $l(g_i) = l(g_k) = l$, *and* $l(g_i g_k) = l$, *then* $g_i = t_\sigma$ *and* $g_k = t'_\sigma$ *or* $= f_{\sigma,\tau}$ *belong to the same set* ϕ_σ.

Proof. Let us assume that g_i and g_k belong to different sets ϕ_{σ_i} and ϕ_{σ_k}.

If $\sigma_i < \sigma_k$, we put again $f = g_i g_k$. Then $l(f) = l$; moreover $l(f^{-1} g_i) = l(g_k^{-1}) = l$, and as g_i belongs to ϕ_{σ_i}, f belongs to $\mathfrak{R}_{\sigma_i+1}$, by 3. 12. But so does g_i, and therefore also $g_k = g_i^{-1} f$, contrary to the assumption that g_k belongs to a later set than g_i.

If $\sigma_i > \sigma_k$, we put $f = g_i^{-1}$. Since $l(g_i g_k) = l(f^{-1} g_k) = l$, $g_i^{-1} = f$ belongs to ϕ_{σ_k+1}, contrary to the assumption that g_i belongs to a later set than g_k.

Hence, g_i and g_k belong to the same set ϕ_σ. But then, by 3. 11, $l(g_i^{-1} g_k) \leq l$. As also $l(g_i g_k) = l$, g_i is a transform, by 2. 4. I. e. $g_i = t_\sigma$, and then $g_k = t'_\sigma$, or $= f_{\sigma\tau}$, as g_k belongs to the same set.

4. 1-4. 3 together give immediately:

4. 4. *If* g_i *and* g_k $(\neq g_i^{-1})$ *are any two generators of equal length* l, *and if* $l(g_i g_k) \leq l$, *then* g_i *and* g_k *belong to the same set* ϕ_σ, *and* $g_i = t_\sigma$ *or* $= f_{\sigma\tau}^{-1}$, *and* $g_k = t'_\sigma$ *or* $f_{\sigma\tau'}$. $l(g_i g_k) < l$ *if, and only if, both* g_i *and* g_k *are transforms,* $g_i = t_\sigma$, $g_k = t'_\sigma$, *and then* $g_i g_k = S_\sigma U_\alpha S_\sigma^{-1}$ *is a* $\mathfrak{U}$*-transform.*

4. 4 leads us to the corresponding results for a word in more than two generators of equal length l:

4. 5.[6] *If* $P = \prod_1^r g_\rho$ $(r > 2)$, *where* $l(g_\rho) = l$ *for* $\rho = 1, \cdots, r$, *and if no proper partial product of* P *is shorter than* l, *then* $l(P) \leq l$ *if, and only if, all generators* g_ρ *belong to the same set* ϕ_σ, $g_\rho = t_\sigma^{(\rho)}$ *for* $\rho = 2, \cdots, r - 1$, *and* g_1 *and* g_r *are such that* $l(g_1 g_r) \leq l$. $l(P) < l$ *only if either also* g_1 *and* g_r *are transforms, or if* $g_1^{-1} = g_r = f_{\sigma\tau}$. *In both cases* P *is a* $\mathfrak{U}$*-transform.*

Proof. Since the product $g_\rho g_{\rho+1}$ of any two successive generators is at least of length l, it follows that in every such product the first $[\frac{1}{2}l]$ factors of g_ρ and the last $[\frac{1}{2}l]$ factors of $g_{\rho+1}$ are not touched by cancellation and amalgamation, and if l is odd, the central factors at most amalgamate. In particular, in the whole word P, the first $[\frac{1}{2}l]$ factors of g_1 and the last $[\frac{1}{2}l]$ factors of g_r remain untouched. If P is to have no more than l factors in all, it follows, that in every product $g_\rho g_{\rho+1}$ at least $[\frac{1}{2}l]$ factors must cancel, and if l is odd, the central factors must amalgamate, i. e. $l(g_\rho g_{\rho+1}) \leq l$ for $\rho = 1, \cdots, r - 1$. Hence, by 4. 4, the first part of 4. 5 follows; and we have

---

[6] Cf. Neumann, [3], p. 18.

$P = g_1 \prod_{2}^{r-1} g_\rho \cdot g_r$ where $g_1 = t_\sigma^{(1)}$ or $= f_{\sigma\tau_1}^{-1}$, and $g_r = t_\sigma^{(r)}$ or $= f_{\sigma,\tau_r}$.

If now $l(P) < l$, then neither of the two generators g_1 and g_r can be earlier than the other. For let g_1, say, be earlier than g_r. As $l(P) < l$, P can be expressed by generators shorter than l, i. e. generators which are earlier than all generators of ϕ_σ. Hence P belongs to $\mathfrak{K}_\sigma$, and, a fortiori, to the group generated by $\mathfrak{K}_\sigma$ and all generators of ϕ_σ which are not later than g_1. But to this group also the partial product $g_1 \prod_{2}^{r-1} t_\sigma^{(\rho)}$ belongs, and therefore also g_r, contrary to the assumption that g_r is later than g_1.

But if neither of the two generators is earlier than the other, then either both are transforms, or one is the inverse of the other, and then it follows from the first part, that $g_r = f_{\sigma\tau} = g_1^{-1}$. I. e. if $l(P) < l$, then either $P = \prod_{1}^{r} t_\sigma^{(\rho)} = S_\sigma U_\alpha S_\sigma^{-1}$ or $P = f_{\sigma\tau}^{-1} \prod_{2}^{r-1} t_\sigma^{(\rho)} \cdot f_{\sigma\tau} = T_{\sigma\tau}^{-1} U_\alpha T_{\sigma\tau}$.

5. Next we investigate the length of the product of a single generator g of length l with a word in generators which are all shorter than l.

5. 0.[7] *If* $P = \prod_{1}^{r} g_\rho$, *and* $\operatorname{Max} l(g_\rho) = l_1$, *and if* g *is a generator of length* $l > l_1$, *then* $l(gP) \geq l(g)$ *and* $l(Pg) \geq l(g)$.

Proof. This is an easy consequence of 4. 0: Let g belong to ϕ_σ. All the generators g_ρ, being shorter than g, belong to earlier sets, hence P belongs to $\mathfrak{K}_\sigma$. If also gP, say, were shorter than g, this also would be expressable by generators of earlier sets than ϕ_σ, i. e. gP, and therefore g, would belong to $\mathfrak{K}_\sigma$, contrary to the fact that g belongs to ϕ_σ.

5. 1. *If, under the same assumptions as in* 5. 0, $l(P) > l_1$, *then* $l(gP) > l(g)$ *and* $l(Pg) > l(g)$; *i. e.* gP *or* Pg *can be of the same length as* g *only if* P *is of the same length as the longest generator involved in its representation.*

Proof. Let us assume that, e. g. $l(gP) = l(g)$, and $l(P) = \lambda > l_1$. If $\lambda = 2\kappa$ is even, then by 1. 12 exactly κ factors of P cancel against g, and no amalgamations can take place. If $\lambda = 2\kappa + 1$ is odd, exactly κ factors of P cancel against g, and one, the central factor of P, amalgamates.

In the first case, $\lambda = 2\kappa$, let $P_2 = \prod_{1}^{r_2} g_\rho$ $(r_2 \leq r)$ be the earliest partial product of P such that P_2 cancels κ factors of g in the product gP_2, while

---

[7] Cf. Neumann, [3], p. 14.

$P_1 = \prod_1^{r_2-1} g_\rho$ cancels less than κ factors in the product gP_1. Such a partial product exists, as by assumption P cancels κ factors of g. Since P_1 cancels $\mu < \kappa$ factors of g and possibly amalgamates one in the product gP_1, but $P_2 = P_1 g_{r_2}$ cancels κ factors of g, the remaining ν factors of P_1 which are not affected by cancellation and amalgamation in the product gP_1, must cancel against g_{r_2}, so that the following factors of g_{r_2} can cancel another $\kappa - \mu$ factors of g in the product $gP_1 g_{r_2}$. Since by 5.0 $l(gP_1) \geqq l(g)$, we have by 1.12: $\nu \geqq \mu$.

In the product $gP_2 = gP_1 g_{r_2}$, altogether $\nu + (\kappa - \mu)$ factors of g_{r_2} have been cancelled. But since also $l(gP_2) \geqq l(g)$, at least κ factors of P_2 are left untouched in gP_2; i. e., since P_1 has been disposed of between g and g_{r_2}, at least κ factors of g_{r_2} are left untouched in this product, so that

$$\begin{aligned} l(g_{r_2}) &\geqq \nu + (\kappa - \mu) + \kappa \\ &\geqq \mu + \kappa - \mu + \kappa = 2\kappa = \lambda > l_1, \end{aligned}$$

against the assumption that all generators g_ρ are of length $\leqq l_1$.

In the second case, where $\lambda = 2\kappa + 1$, let $P_2 = \prod_1^{r_2} g_\rho$ be the earliest partial product which cancels κ and amalgamates one factor of g in the product gP_2, while $P_1 = \prod_1^{r_2-1} g_\rho$ either cancels less than κ factors of g, or cancels κ factors, but then does not amalgamate a further factor of g in the product $P_1 g$. If P_1 cancels μ factors of g, we have $\mu \leqq \kappa$. If ν factors of P_1 are left unaffected in the product gP_1, we have again, because of $l(gP_1) \geqq l(g)$, $\nu \geqq \mu$.

These last ν factors of P_1 must cancel against g_{r_2}, so that then g_{r_2} can cancel another $\kappa - \mu$ factors of g, and amalgamate one, in the product $gP_1 g_{r_2} = gP_2$. Because of $l(gP_2) \geqq l(g)$, there must be at least κ factors left of g_{r_2} which are not touched by cancellation and amalgamation, so that

$$\begin{aligned} l(g_{r_2}) &\geqq \nu + (\kappa - \mu) + 1 + \kappa \\ &\geqq \mu + \kappa - \mu + 1 + \kappa = 2\kappa + 1 = \lambda > l_1, \end{aligned}$$

which is the same contradiction as before. Hence $l(gP) = l(g)$ is impossible if $l(P) > l_1$. If $l(gP) > l(g)$, but $l(Pg) = l(g)$, the proof runs similarly, and can be omitted.

We call a word in generators g_ρ of length $l(g_\rho) \leqq l$ $(\rho = 1, \cdots, r)$, which cannot be expressed by generators which are all shorter than l, a 'prime word,'[8] if it is itself of length l, i. e. of the same length as the longest generator involved in its representation. We denote prime words by $p, p', \cdots$. Then 5.1 becomes:

5.2. *If* $P = \prod_1^r g_\rho$ *is a word in generators of length* $l(g_\rho) \leq l_1$ $(\rho = 1, \cdots, r)$, *and if there exists a generator* g *of length* $l(g) > l_1$ *so that either* $l(Pg) = l(g)$ *or* $l(gP) = l(g)$, *then* $P = p$ *is a prime word of length* $l(p) \leq l_1$.

Finally we prove:

5.3.[9] *If* $P = \prod_1^r g_\rho$, $l(g_\rho) \leq l_1$ $(\rho = 1, \cdots, r)$, *and if* g *is a generator of length* $l > l_1$, *then* $l(Pg) > l(P)$ *and* $l(gP) > l(P)$.

Proof. Again it will be sufficient to prove one of the inequalities only, the proof of the other one being similar.

Let us assume that $l(gP) \leq l(P)$, and let $P_1 = \prod_1^{r_1} g_\rho$ be the latest partial product of P such that $l(gP_1) > l(P_1)$; but if $P_2 = \prod_1^{r_1+1} g_\rho$, then $l(gP_2) \leq l(P_2)$. Such a partial product P_1 exists, since by 5.0 $l(gg_1) \geq l(g) > l(g_1)$.

Since $l(gP_1) > l(P_1)$, by 1.12 more factors of g remain untouched than cancel against P_1, i. e. if l is odd, at least the first half and the central factor of g remain untouched, and if l is even, at least the first half of g remains untouched, in the product gP_1. If we write $\{\frac{1}{2}l\} = l - [\frac{1}{2}l]$,[10] this means, that at least the first $\{\frac{1}{2}l\}$ factors of gP_1 are simply the factors of g, and if $l(gP_1) = \{\frac{1}{2}l\} + \rho$, we have, because of $l(gP_1) \geq l(g)$, $\rho \geq [\frac{1}{2}l]$.

Since $l(gP_2) \leq l(P_2)$, P_2 cancels in this product $gP_2 = gP_1 g_{r_1+1}$ at least the second half of g, and if l is odd, it also amalgamates or cancels the central factor of g. Hence, in this product, g_{r_1+1} must cancel at least the last ρ factors of gP_1, and, if l is odd, cancel or amalagamate one more. If in the product $gP_2 = (gP_1)g_{r_1+1}$ exactly σ factors cancel, we have $\sigma \geq \rho$, and by 1.10:

$$l(gP_2) = l(gP_1) + l(g_{r_1+1}) - 2\sigma - \epsilon,$$

where from the preceding discussion we know that $\epsilon = 1$, if $\sigma = \rho$ and l is odd,

---

[8] Cf. Kurosch, [1], p. 652.

[9] Cf. Neumann, [3], p. 16.

[10] While $[\frac{1}{2}l]$ is the length of the first or second half of an element of length l, $\{\frac{1}{2}l\}$ is the length of first or second half plus central factor.

i. e. $\{\frac{1}{2}l\} = [\frac{1}{2}l] + 1$, but if $\epsilon = 0$ and $\sigma = \rho$, then l must be even, i. e. $\{\frac{1}{2}l\} = [\frac{1}{2}l]$. It follows:

$$\begin{aligned} l(gP_2) &= \{\tfrac{1}{2}l\} + \rho + l(g_{r_1+1}) - 2\sigma - \epsilon \\ &\leq [\tfrac{1}{2}l] + 1 + \rho + l(g_{r_1+1}) - 2\rho - 1 \\ &\leq \rho + 1 + \rho + l(g_{r_1+1}) - 2\rho - 1 \\ &= l(g_{r_1+1}) < l(g) \end{aligned}$$

contrary to 5. 0.

6. We deal now with words of the form g_iPg_k, where g_i and g_k are generators of equal length l, and P is a word in shorter generators.

It is here that the difference between the generalized free product and the ordinary free product becomes apparent. In the ordinary free product, every word of this form is longer than l, i. e. longer than the generators g_i and g_k,[11] whereas in our case its length can be less than, or equal to l. This difference alone is responsible for the existence of non-trivial relations between our generators, as will be seen in the following paragraphs. In this paragraph we shall have to find out under what conditions g_iPg_k can be shorter than g_i and g_k.

6. 0. *If at least one of the inequalities* $l(g_iP) > l = l(g_i)$ *or* $l(Pg_k) > l = l(g_k)$ *holds, then* $l(g_iPg_k) > l$.

Proof. If, e. g. $l(g_iP) > l$, and if $l(P) = \lambda$, then by 1. 12, less than $[\frac{1}{2}\lambda]$ factors of P cancel against g_i, and at most $[\frac{1}{2}\lambda]$ factors cancel and amalgamate. Besides, by 5. 3, $l(g_iP) > l(P)$; hence at least the first $\{\frac{1}{2}l\}$ factors of g_i remain untouched in the product g_iP. And as also $l(Pg_k) > l(P)$, also at least the last $\{\frac{1}{2}l\}$ factors of g_k remain untouched in the product Pg_k. Moreover, $l(Pg_k) \geq l(g_k)$ by 5. 1, which means that at most $[\frac{1}{2}\lambda]$ factors of P cancel against g_k, and at most $\{\frac{1}{2}\lambda\}$ factors cancel and amalgamate. In the whole product g_iPg_k, at least one (possibly amalgamated) factor of P remains therefore between the first $\{\frac{1}{2}l\}$ factors of g_i and the last $\{\frac{1}{2}l\}$ factors of g_k, i. e. $l(g_iPg_k) \geq \{\frac{1}{2}l\} + 1 + \{\frac{1}{2}l\} > l$.

6. 1. *If* $l(g_iPg_k) \leq l$, *then* $P = p$ *is a prime word of length* $l(p) < l$, *and* g_i *and* g_k *belong to the same set* ϕ_σ *so that, moreover,* $l(g_ig_k) \leq l$.

Proof. From $l(g_iPg_k) \leq l$, together with 6. 0 and 5. 0, it follows that $l(g_iP) = l = l(Pg_k)$, hence, by 5. 2, $P = p$ is a prime word of length $l(p) < l$.

---

[11] Cf., e. g., Neumann, [3], p. 17.

Let us assume that g_i and g_k belong to different sets ϕ_{σ_i} and ϕ_{σ_k}, and, without loss of generality, that $\sigma_i < \sigma_k$.

g_i is either of the form $g_i = t_{\sigma_i}$ or $= f_{\sigma_i\tau}$, or $g_i^{-1} = f_{\sigma_i\tau}$.

In the first case, we put $f = g_i p g_k$. Then either $l(f) < l$, in which case f belongs to $\mathfrak{R}_{\sigma_i}$. As p is also shorter than l, p also belongs to $\mathfrak{R}_{\sigma_i}$. Hence $g_k = p^{-1} g_i^{-1} f$ belongs to $\mathfrak{R}_{\sigma_i+1}$, contrary to the assumption that g_k belongs to a set later than ϕ_{σ_i}.

Or $l(g_i p g_k) = l(f) = l$. But then $l(f^{-1} g_i) = l(g_i^{-1} f) = l(p g_k) = l$, i. e. by 3. 12, f belongs to $\mathfrak{R}_{\sigma_i+1}$; so does $g_i p$, and therefore also g_k, which is the same contradiction as before.

In the other case, where $g_i^{-1} = f_{\sigma_i\tau}$, we put $f = p g_k$. Then $l(f) = l$ and $l(f^{-1} g_i^{-1}) = l$, hence again f belongs to $\mathfrak{R}_{\sigma_i+1}$, and therefore also g_k does, in contradiction to the assumption that g_k belongs to a later set than g_i.

Therefore, g_i and g_k must belong to the same set ϕ_σ. It remains to be proved that $l(g_i g_k) \leq l$. That $l(g_i g_k) > l$ can happen only in the following two ways (but for trivial variations):

$$\begin{array}{lll} \text{either } g_i = t_\sigma \text{ or } = f_{\sigma\tau_1}^{-1} & \text{and} & g_k = f_{\sigma\tau_2}^{-1}, \\ \text{or} \quad\ g_i = f_{\sigma\tau_1} & \text{and} & g_k = f_{\sigma\tau_2}^{-1}. \end{array}$$

In the first case, with $f = p g_k$, we have $l(f) = l$, $l(f^{-1} g_i^{-1}) \leq l$, and $l(f g_k^{-1}) \leq l$, where both, g_i^{-1} and g_k^{-1}, are generators of ϕ_σ beginning with S_σ. Hence, by 3. 12, $f = t'_\sigma$ is itself a generator of ϕ_σ. But then $g_k = p^{-1} t'_\sigma$ cannot be a generator later than t'_σ, contrary to $g_k = f_{\sigma\tau_2}^{-1}$. In the second case, if $g_i \neq g_k^{-1}$, one of them is earlier than the other, g_i earlier than g_k, say. We put $f = g_i p g_k$. If f were shorter than l, f would lie in $\mathfrak{R}_\sigma$, hence $g_k = p^{-1} g_i^{-1} f$ were not later than g_i. Hence f is of length l. But also $l(g_i^{-1} f) = l(f^{-1} g_i) = l$ and $l(f g_k^{-1}) = l$, where both g_i and g_k^{-1} are generators of ϕ_σ beginning with S_σ. It follows again from 3. 12, that $f = t_\sigma$, and $g_k = p^{-1} g_i^{-1} f = p^{-1} g_i^{-1} t_\sigma$ is expressible by generators not later than g_i, which is a contradiction.

Therefore, in fact, $l(g_i g_k) \leq l$; which completes the proof of 6. 1.

Under the assumptions of 6. 1, we have $l(g_i p) = l(p g_k) = l$ and $l(g_i g_k) \leq l$, i. e. $l(p^{-1} g_i^{-1}) = l(g_i)$ and $l(g_i g_k) \leq l(g_k)$, hence by 2. 11 $l(p^{-1} g_k) \leq l(g_k)$. As also $l(p g_k) = l(g_k)$, it follows from 2. 4 that p is a transform. Closer examination shows that p is a $\mathfrak{U}$-transform. This is seen as follows:

Because of $l(g_i g_k) \leq l$, we may write representations 1. 0 of g_i and g_k,

$g_i = S(g_i)C(g_i)T(g_i)$ and $g_k = S(g_k)C(g_k)T(g_k)$, in such a way that $T(g_i) \equiv S(g_k)^{-1}$. Besides $C(g_i)$ and $C(g_k)$ belong to the same group.

By 5.3, $l(pg_k) > l(p)$, so that $C(g_k)T(g_k)$ is unaffected in the product pg_k. But we also have $l(pg_k) = l = l(g_k)$, i. e. the first half of pg_k is simply the product of p and $S(g_k)$. Therefore, a representation 1.0 of pg_k may be written in the form $pg_k = (pS(g_k)) \cdot C(g_k)T(g_k)$. Now the whole product $g_i p g_k$ is a product of two factors of equal length l (viz. g_i and pg_k), and itself at most of length l; hence $T(g_i)$ cancels against $S(pg_k) = pS(g_k)$, and, if l is odd, the central factors $C(g_i)$ and $C(pg_k) = C(g_k)$ cancel or amalgamate, i. e. the product $T(g_i)pS(g_k)$ is an element of $\mathfrak{U}_\alpha$ if the central factors of g_i and g_k belong to $\mathfrak{G}_\alpha - \mathfrak{U}_\alpha$. Hence, for any l, we have

$$T(g_i)pS(g_k) = U \text{ in } \mathfrak{U},$$

i. e.

$$p = T(g_i)^{-1}US(g_k)^{-1} = S(g_k)US(g_k)^{-1}.$$

If l is odd, $U = U_\alpha$, and $p = S(g_k)U_\alpha S(g_k)^{-1}$ is a $\mathfrak{U}$-transform; if l is even, at least one factor each of $S(g_k)$ and $S(g_k)^{-1}$ must cancel or amalgamate in $p = S(g_k)US(g_k)^{-1}$, as $l(p) < l$; therefore $U = U_\beta$ where $\mathfrak{G}_\beta$ is the group of the last factor of $S(g_k)$, and $p = S(g_k)U_\beta S(g_k)^{-1}$ is again a $\mathfrak{U}$-transform.

We call a prime word which is a $\mathfrak{U}$-transform, a "prime $\mathfrak{U}$-transform." Prime $\mathfrak{U}$-transforms will be denoted by $\pi, \pi', \cdots$. Then we have shown:

6.2. *If g_i and g_k are two generators of equal length l, P is a word in shorter generators, then $l(g_iPg_k) \leq l$ if, and only if, g_i and g_k belong to the same set ϕ_σ and $l(g_ig_k) \leq l$, and P is a prime $\mathfrak{U}$-transform, of length $< l$,*

$$P = \pi = S(g_k)\mathfrak{U}_\alpha S(g_k)^{-1},$$

where α indicates the group of the central factors of g_i and g_k if l is odd, and the group of the last factor of $S(g_k)$, if l is even.

If in 6.2, $g_i\pi g_k$ is actually shorter than l, then $g_i\pi g_k = P'$ is itself a word in generators which are all shorter than l, moreover $g_i^{-1}P'g_k^{-1} = \pi$ is shorter than l. Hence, 6.2 applied to $g_i^{-1}P'g_k^{-1}$ gives that also $P' = \pi'$ is a prime $\mathfrak{U}$-transform, and also $l(g_i^{-1}g_k^{-1}) \leq l$. Using 4.2 and 4.3, we get

6.30. *If, under the same assumptions as in 6.2, we have $l(g_iPg_k) < l$ then either $g_i^{-1} = g_k = f_{\sigma\tau}^{\epsilon}$ $(\epsilon = \pm 1)$, and then $g_iPg_k = f_{\sigma\tau}^{-\epsilon}\pi f_{\sigma\tau}^{\epsilon} = \pi'$, where both, π and π', are prime $\mathfrak{U}$-transforms shorter than l, or $g_i = t_\sigma$, $g_k = t'_\sigma$, and then $g_iPg_k = t_\sigma\pi t'_\sigma = \pi'$, where π and π' are again prime $\mathfrak{U}$-transforms shorter than l.*

For later reference, we give the explicit representation of π and π', which is easily checked by means of 6. 30 and the preceding discussion:

6. 31. (i) *l odd:*

If $g_i^{-1} = g = f_{\sigma\tau} = S_\sigma C_\sigma T_{\sigma\tau}$, then $f_{\sigma\tau}^{-1}\pi f_{\sigma\tau} = \pi'$, where $\pi = S_\sigma U_\alpha S_\sigma^{-1}$, $\pi' = T_{\sigma\tau}^{-1}U'_\alpha T_{\sigma\tau}$, with $C_{\sigma\tau}^{-1}U_\alpha C_{\sigma\tau} = U'_\alpha$; if $g_i = t_\sigma = S_\sigma C_\sigma S_\sigma^{-1}$, $g_k = t'_\sigma = S_\sigma C'_\sigma S_\sigma^{-1}$, then $\pi = S_\sigma U_\alpha S_\sigma^{-1}$, $\pi' = S_\sigma U'_\alpha S_\sigma^{-1}$, with $C_\sigma U_\alpha C'_\sigma = U'_\alpha$.

(ii) *l even:*

If $g_i^{-1} = g_k = f_{\sigma\tau} = S_\sigma U T_{\sigma\tau}$, where the last factor of S_σ lies in $\mathfrak{G}_\beta - \mathfrak{U}_\beta$, and the first factor of $T_{\sigma\tau}$ lies in $\mathfrak{G}_\gamma - \mathfrak{U}_\gamma$, then $f_{\sigma\tau}^{-1}\pi f_{\sigma\tau} = \pi'$, where $\pi = S_\sigma U_\beta S_\sigma^{-1}$, $\pi' = T_{\sigma\tau}^{-1}U_\gamma T_{\sigma\tau}$ with $U^{-1}U_\beta U = U_\gamma$; if $g_i = t_\sigma = S_\sigma U S_\sigma^{-1}$, $g_k = t'_\sigma = S_\sigma U' S_\sigma^{-1}$, then $\pi = S_\sigma U_\beta S_\sigma^{-1}$, $\pi' = S_\sigma U'_\beta S_\sigma^{-1}$, with $U U_\beta U' = U'_\beta$.

7. After these preparations, we are able to consider arbitrary words in our generators. Our aim is to obtain a relation between the length of the word as element of $\mathfrak{H}$, and the length of the longest generator involved in its representation. But if any word

$$7.0 \qquad P = \prod_{\rho=1}^{r} g_\rho, \qquad \operatorname{Max} l(g_\rho) = l,$$

is given, it may yet be that P can also be expressed by shorter generators only—examples for this are the situations of 4. 5 and 6. 3—and then we cannot expect to get an estimate of the length of P in terms of the length of the longest generator in this particular representation.

The lemmas 4. 5, 6. 2, and 6. 3 provide some means to reduce, possibly, the number of longest generators in the representation 7. 0. In this and the following paragraph it will be shown that these three lemmas suitably applied represent *all* the possibilities of disposing of generators of maximal length in a word 7. 0.

We refer to a process which reduces the number of longest generators in the word 7. 0 as a 'reduction' of the word 7. 0.

It will be useful to list all the different types of reductions based on 4. 5, 6. 2, and 6. 3. To this end, we write 7. 0 in the following form:

$$7.1 \qquad P = \prod_{\rho=1}^{r} g_\rho = \prod_{\nu=1}^{n} P_\nu \qquad (r \geq n),$$

where every factor P_ν is either a generator of length l, or the partial product of all the shorter generators preceding the first, or following the last generator of length l, or between two successive generators of length l.

Moreover, we assume that in 7. 1 a generator of length l which is of the form $P_\nu = t_\sigma$ is neither followed nor preceded by a prime $\mathfrak{U}$-transform $P_{\nu\pm1} = \pi = S_\sigma U_\alpha S_\sigma^{-1}$. For in such a case, the products $t_\sigma\pi$ or πt_σ are both of length l, and therefore themselves generators t'_σ. Hence the products $P_\nu P_{\nu+1} = t_\sigma\pi$ or $P_{\nu-1}P_\nu = \pi t_\sigma$ can be replaced by the one factor t'_σ.

We begin the list of reductions with the obvious one, viz. replacing a number $r \geqq 2$ of successive generators of the form $t_\sigma^{(1)}, t_\sigma^{(2)}, \cdots$ by their product. This is either of length l and then itself a generator t_σ, or it is shorter than l and then, by 4. 5, a $\mathfrak{U}$-transform π, as e. g. for every generator t_σ we have $l(\pi t_\sigma) = l(t_\sigma)$. The first reduction is therefore:

(i) $r \geqq 2$ successive factors $P_{\nu+\rho} = t_\sigma^{(\rho)} = S_\sigma C_\sigma^{(\rho)} S_\sigma^{-1}$ $(\rho = 1, \cdots, r)$ are replaced by their product $\prod_{\rho=1}^{r} P_{\nu+\rho} = t_\sigma$ or $= \pi = S_\sigma U_\alpha S_\sigma^{-1}$.

If three successive factors are all generators of length l, but not all of the form t_σ, then by 4. 5 their product can be shorter than l only if $P_{\nu-1} = f_{\sigma\tau}^{-1}$, $P_\nu = t_\sigma$, $P_{\nu+1} = f_{\sigma\tau}$, and then

$$P_{\nu-1}P_\nu P_{\nu+1} = \pi = T_{\sigma\tau}^{-1} U_\alpha T_{\sigma\tau}$$

is a prime $\mathfrak{U}$-transform, by 6. 2, as $f_{\sigma\tau}\pi_{\sigma\tau}^{-1} = t_\sigma$ is of length l. But here also t_σ is a $\mathfrak{U}$-transform (hence a prime $\mathfrak{U}$-transform), for

$$t_\sigma = f_{\sigma\tau}(T_{\sigma\tau}^{-1} U_\alpha T_{\sigma\tau}) f_{\sigma\tau}^{-1} = S_\sigma(C_{\sigma\tau} U_\alpha C_{\sigma\tau}^{-1}) S_\sigma^{-1}$$

i. e. $C_\sigma = C(t_\sigma) = C_{\sigma\tau} U_\alpha C_{\sigma\tau}^{-1}$.

The second reduction is therefore:

(ii) Three successive generators $f_{\sigma\tau}^{-1}$, t_σ, $f_{\sigma\tau}$ are replaced by their product, a prime $\mathfrak{U}$-transform:

$$f_{\sigma\tau}^{-1} t_\sigma f_{\sigma\tau} = \pi = T_{\sigma\tau}^{-1} U_\alpha T_{\sigma\tau} \text{ with } U_\alpha = C_{\sigma\tau}^{-1} C_\sigma C_{\sigma\tau}.$$

This same relation gives rise to another reduction if we write it in the form

$$t_\sigma f_{\sigma\tau} = f_{\sigma\tau}\pi \text{ or } f_{\sigma\tau}^{-1} t_\sigma = \pi f_{\sigma\tau}^{-1},$$

which replaces two generators of length l by only one and a word in shorter generators. And the same reduction is still applicable even if not t_σ itself satisfies the above relation, as long as it can be written in the forms $t_\sigma = \pi_1 t'_\sigma$

or $t_\sigma = t'_\sigma \pi_1$ respectively, where $\pi_1 = S_\sigma U_{a_1} S_\sigma^{-1}$, and t'_σ satisfies the above relation. Therefore:

(iii) Two successive generators of length l, $P_\nu = t_\sigma$ and $P_{\nu+1} = f_{\sigma\tau}$, are replaced by their product

$$P_\nu P_{\nu+1} = t_\sigma f_{\sigma\tau} = \pi_1 f_{\sigma\tau} \pi \text{ with } \pi_1 = S_\sigma U_{a_1} S_\sigma^{-1},\ \pi = T_{\sigma\tau}^{-1} U_a T_{\sigma\tau};$$

and correspondingly $P_\nu = f_{\sigma\tau}^{-1}$ and $P_{\nu+1} = t_\sigma$ are replaced by

$$P_\nu P_{\nu+1} = f_{\sigma\tau}^{-1} t_\sigma = \pi f_{\sigma\tau}^{-1} \pi_1 \text{ with } \pi \text{ and } \pi_1 \text{ as before.}$$

Combining the two reductions (iii) leads to a reduction of the form:

(iv) Three successive generators of length l, $P_{\nu-1} = f_{\sigma\tau}^{-1}$, $P_\nu = t_\sigma$ and $P_{\nu+1} = f_{\sigma\tau_2}$ are replaced by their product

$$P_{\nu-1} P_\nu P_{\nu+1} = f_{\sigma\tau_1}^{-1} t_\sigma f_{\sigma\tau_2} = \pi_1 f_{\sigma\tau_1}^{-1} \pi f_{\sigma\tau_2} \pi_2,$$

where $\pi_1 = T_{\sigma\tau_1}^{-1} U_\beta T_{\sigma\tau_1}$, $\pi = S_\sigma U_a S_\sigma^{-1}$, $\pi_2 = T_{\sigma\tau_2}^{-1} U_\gamma T_{\sigma\tau_2}$.

By 6. 2, a product of the form $g_i P g_k$ will be itself a generator of length l only if $g_i = f_{\sigma\tau}$, $g_k = f_{\sigma\tau}^{-1}$, $P = \pi = T_{\sigma\tau}^{-1} U_\beta T_{\sigma\tau}$, and then $g_i \pi g_k = t_\sigma$, where t_σ is a prime $\mathfrak{U}$-transform as in (ii). As in this form two generators are replaced by one, we list it as a separate reduction:

(v) Three successive factors of the form $P_{\nu-1} = f_{\sigma\tau}$, $P_\nu = \pi = T_{\sigma\tau}^{-1} U_\beta T_{\sigma\tau}$, $P_{\nu+1} = f_{\sigma\tau}^{-1}$ are replaced by their product

$$P_{\nu-1} P_\nu P_{\nu+1} = f_{\sigma\tau} \pi f_{\sigma\tau}^{-1} = t_\sigma = S_\sigma C_\sigma S_\sigma^{-1} \text{ with } C_\sigma = C_{\sigma\tau} U_\beta C_{\sigma\tau}^{-1}.$$

Finally, a product $g_i P g_k$ which is shorter than l, can be replaced by a word in shorter generators only. By 6. 3, and our convention that a generator t_σ is neither followed nor preceded by a factor $P_\nu = \pi = S_\sigma U_a S_\sigma^{-1}$, this leads to the following reduction:

(vi) Three successive factors of the form $P_{\nu-1} = f_{\sigma\tau}^{-\epsilon}$, $P_\nu = \pi = S(f_{\sigma\tau}^{\epsilon}) U_a S^{-1}(f_{\sigma\tau}^{\epsilon})$, and $P_{\nu+1} = f_{\sigma\tau}^{\epsilon}$ $(\epsilon = \pm 1)$ are replaced by their product:

$$f_{\sigma\tau}^{-\epsilon} \pi f_{\sigma\tau}^{\epsilon} = \pi_1 = T^{-1}(f_{\sigma\tau}^{\epsilon}) U_\beta T(f_{\sigma\tau}^{\epsilon}).$$

The process of reducing 7. 1 by means of the reductions (i)-(vi) will come to an end either before all generators of length l in 7. 1 are eliminated, or when all generators of length l have disappeared. In that case we are left with a representation of P by means of generators of length $\leq l_1 < l$ only. We repeat the process with respect to the now longest generators of length l_1, and so on. After a finite number of steps, no further reduction will be possible. A representation of P thus obtained is called a reduced representation of P, or simply a reduced word P. From now on we assume that 7. 1 is a reduced word, where the length of the longest generator is again l.

The following properties of reduced words follow immediately:

7. 2. *With* $P = \prod_1^n P_\nu$ *also the formal inverse* $P^{-1} = \prod_n^1 P_\nu^{-1}$ *is a reduced word.*

7. 3. *Every partial product* $\prod_{n_0}^{n_1} P_\nu$ $(n \geq n_1 > n_0 \geq 1)$ *of a reduced word* $P = \prod_1^n P_\nu$, *which contains at least one factor* P_ν $(n_0 \leq \nu \leq n_1)$ *which is a generator of length* l, *is itself a reduced word.*

7. 4. *If* Q_1 *and* Q_2 *are words in generators which are shorter than* l, *and* $P = \prod_1^n P_\nu$ *a reduced word, then the reduced representation of* $Q_1PQ_2 = Q_1 \prod_1^n P_\nu \cdot Q_2$ *contains the same number of generators of length* l *as* P.

Proof. The reduced representation of Q_1PQ_2 is obtained as follows: if P_1 is a word in shorter generators, combine Q_1 and P_1 into one factor $Q_1P_1 = P'_1$. The only case in which this may give rise to a further change is where P'_1 is of the form $\pi = S_\sigma U_a S_\sigma^{-1}$ and P_2 is a generator of the form t_σ. In that case we have to combine P'_1 and P_2 into one factor t'_σ. This does not change the number of longest generators in P, as it cannot cause a reduction; any reduction caused by replacing t_σ by t'_σ would have to be of the type (iii)), but then already the original generator t_σ would have given rise to this reduction, contrary to the assumption that P was reduced. A similar argument shows that Q_2 has no effect on the number of longest generators in P.

We are now going to prove the decisive property of reduced words:

7. 5.[12] *If* $P = \prod_1^n P_\nu$ *is a reduced word, and if the length of its longest generator is* l, *then* $l(P) \geq l$.

Proof. We bring this proof in several steps.

1. By 4. 1, 5. 0, and 5. 3, any two successive factors P_ν and $P_{\nu+1}$ cancel at most half of each other; and if P_ν is a word in shorter generators, hence

---

[12] Cf. Neumann, [3], p. 19.

$P_{\nu-1}$ and $P_{\nu+1}$ generators of length l, P_ν cancels definitely less than the first half $S(P_{\nu+1})$ of $P_{\nu+1}$, and less than the second half $T(P_{\nu-1})$ of $P_{\nu-1}$. In particular, if P_1 is a word in shorter generators, hence P_2 the first generator of length l, then P_1 affects less than the first half $S(P_2)$ of P_2, and the product $P_1 \cdot S(P_2)$ is at least of the same length $[\frac{1}{2}l]$ as $S(P_2)$. Correspondingly, if P_n is a word in shorter generators, hence P_{n-1} the last generator of length l, then less than the second half $T(P_{n-1})$ is affected in the product $P_{n-1}P_n$, and $T(P_{n-1}) \cdot P_n$ is at least of length $[\frac{1}{2}l]$.

2. If the first or second half of a generator P_ν of length l is cancelled completely by the factors preceding, or following, P_ν then the second or first half respectively of the generator of length l preceding, or following, P_ν is thereby cancelled as well. For if, e. g., $S(P_\nu)$ is cancelled, and if $P_{\nu-1}$ is also a generator of length l, then, as any two successive factors cancel at most half of each other, $S(P_\nu)$ must be cancelled by $T(P_{\nu-1})$. But if $P_{\nu-1}$ is a word in shorter generators, hence $P_{\nu-2}$ the generator of length l preceding P_ν, then $P_{\nu-1}$ does not cancel the whole of $S(P_\nu)$, hence $P_{\nu-1}$ must have been disposed of between the generators $P_{\nu-2}$ and P_ν, so that then the former can cancel the rest of $S(P_\nu)$. But then we have the position of 6. 2, i. e. $T(P_{\nu-2}) = S(P_\nu)^{-1}$, and $P_{\nu-1} = \pi = S(P_\nu)U_\alpha S(P_\nu)^{-1}$, and both, $S(P_\nu)$ and $T(P_{\nu-2})$, cancel completely.

3. If P_ν is a generator of length l of the form $P_\nu = f_{\sigma\tau}^{\pm 1}$, then if l is even, the half $T_{\sigma\tau}^{\pm 1}$ cannot be cancelled completely, and if l is odd, $C_{\sigma\tau}^{\pm 1}$ cannot be altered by the factors following P_ν (if $P_\nu = f_{\sigma\tau}$) or preceding P_ν (if $P_\nu = f_{\sigma\tau}^{-1}$). For let, e. g., $P_\nu = f_{\sigma\tau}$. By 4. 2 and 4. 3, there is no other generator g, i. e. no generator $g \neq f_{\sigma\tau}^{-1}$, of length l such that $l(f_{\sigma\tau}g) \leq l$, or $l(f_{\sigma\tau}g^{-1}) \leq l$. Hence, if the second half $T_{\sigma\tau}$ of $f_{\sigma\tau}$ should cancel completely, and, if l is odd, the central factor $C_{\sigma\tau}$ at least amalgamate, $f_{\sigma\tau}$ must be followed by a product $P_{\nu+1}$ of shorter generators and a generator $P_{\nu+2}$ of length l, which together cancel $T_{\sigma\tau}$, and if l is odd, amalgamate $C_{\sigma\tau}$. But then we have, by 6. 2, $P_{\nu+2} = P_\nu^{-1} = f_{\sigma\tau}^{-1}$ and $P_{\nu+1} = T_{\sigma\tau}^{-1}U_\beta T_{\sigma\tau}$, and P would have been reduced. Similarly it follows in the case that $P_\nu = f_{\sigma\tau}^{-1}$.

4. Now we show: in the whole word P, the first half of the first generator of length l, and the second half of the last generator of length l, are not touched at all by cancellation or amalgamation. And, if l is odd, the central factors of the first and last generators of length l can, in the whole word P, at most amalgamate, but never cancel.

Again it will be sufficient to prove this assertion for the first generator only, the proof for the last generator being analogous. If the first generator of length l is of the form $f_{\sigma\tau}$, we saw under 3. that in the whole product P cancellation and amalgamation can touch at most the second half $T_{\sigma\tau}$, while the first half and the central factor remain untouched.

If the first generator of length l is of the form $f_{\sigma\tau}^{-1}$ or t_{σ}. so that its second half is S_{σ}^{-1}, and if this half is cancelled completely, then, by 2., this is done by the next following generator of length l.

Either this is of the form $f_{\sigma\tau_1}$. Then it follows from 3. that the next following generator of length l can have no influence on what happens to the first. If then the first half of the first generator, or, if l is odd, its central factor, should be touched at all, the partial product $f_{\sigma\tau}^{-1}P_3 f_{\sigma\tau_1}$ (P_3 a word in shorter generators) or $T_\sigma P_3 f_{\sigma\tau_1}$ respectively of P must be of length $\leq l$. If it were $< l$, P would not have been reduced by 6. 3. Hence it is of length l, i. e. if l is odd, the central factor of the first generator amalgamates, but does not cancel, and if l is even, no amalgamation can take place between the first half of the first generator and the second half of the second generator.

Or else the second generator is of the form t'_{σ}, and in order that it cancels the second half of the first generator, a factor between the two would have to be a prime $\mathfrak{U}$-transform of the form $S_\sigma U_a S_\sigma^{-1}$, by 6. 2; but this possibility has been excluded. Hence the first generator of length l is followed immediately by t'_{σ}, and as P is reduced, the first generator can only be of the form $f_{\sigma\tau}^{-1}$. In the product $f_{\sigma\tau}^{-1}t'_{\sigma}$, $S(f_{\sigma\tau}^{-1})$ is not affected at all, and if l is odd, the two central factors amalgamate, but do not cancel. Cancellation, therefore, could only be achieved by the next following generator of length l, which then would have to cancel the second half, and amalgamate (at least) the central factor, of t'_{σ} first. For the same reason as before, no word in shorter generators can occur between the second and third generator of length l, hence the third one must be of the form $f_{\sigma\tau'}$ and we would have a product of the form $f_{\sigma\tau}^{-1}t'_{\sigma}f_{\sigma\tau'}$. As P is reduced, this must be of length $\geq l$, i. e. $S(f_{\sigma\tau}^{-1})$ is not touched, and if l is odd, the central factors at most amalgamate. Moreover, by 3., the factors following $f_{\sigma\tau'}$ cannot alter this position any more.

Now it follows from 1. and 4.:

$$\text{if } l \text{ is even: } l(P) \geq [\tfrac{1}{2}l] + [\tfrac{1}{2}l] = l,$$

$$\text{if } l \text{ is odd: } l(P) \geq [\tfrac{1}{2}l] + 1 + [\tfrac{1}{2}l] = l, \text{ q. e. d.}$$

Now the main result of this paragraph follows easily:

7.6. *If the longest generators in a reduced word P are of length l, then P cannot be expressed by generators which are all shorter than l.*

Proof. Let $P = \prod_1^r g_\rho = \prod_1^n P_\nu$ be the reduced word whose longest generators are of length l. If $P_1 = \prod_1^{r'} g'_\rho$ with $l(g'_\rho) < l$ for all ρ were another representation of the same element of $\mathfrak{H}$, then $P_1^{-1}P = 1$. But P_1 is a word in generators which are all shorter than l, hence, by 7.4, a reduced representation of $P_1^{-1}P$ contains the same number of generators of length l as P. Hence, by 7.5, $l(P_1^{-1}P) \geq l$, contrary to $P_1^{-1}P = 1$.

8. From **7**, in particular from 7.5 and 7.6, we derive a complete description of a system of defining relations for the generators of $\mathfrak{H}$.

If $R = \prod_1^r g_\rho$ with $\operatorname{Max} l(g_\rho) = l$ is a relation between the generators of our system ϕ, i.e. $R = 1$ as element of $\mathfrak{H}$, the reduction of R cannot leave any generators of length l, by 7.5. The same is true for the reduction with respect to the then longest generators, and so on. Hence, after a finite number of steps, the right-hand side will become identically one by means of the reductions (i)-(vi). It follows:

8.0. THEOREM. *Every relation between generators of the system ϕ follows from relations of one of the following two types:*

8.01. $\prod_{\rho=1}^r t_\sigma^{(\rho)} = \pi$ *where* $\pi = S_\sigma U_\alpha S_\sigma^{-1}$ *is a prime* $\mathfrak{U}$*-transform of length* $l(\pi) < l(t_\sigma)$, *or the unit element;*

8.02. $f_{\sigma\tau}^{-1} T f_{\sigma\tau} = \pi_1$ *where* $T = t_\sigma = S_\sigma(C_{\sigma\tau} U_\alpha C_{\sigma\tau}^{-1}) S_\sigma^{-1}$ *or* $T = \pi = S_\sigma U_\alpha S_\sigma^{-1} \neq 1$, *and* $\pi_1 = T_{\sigma\tau}^{-1} U_\beta T_{\sigma\tau}$.

(Here, if $l(t_\sigma)$ or $l(f_{\sigma\tau})$ respectively is odd, α indicates the group of the central factors of the generators of the set ϕ_σ; if this length is even, α indicates the group of the last factor of S_σ and β that of the first factor of $T_{\sigma\tau}$.)

Proof. All that remains to be proved is that the reductions (i)-(vi) follow from relations of the form 8.01 and 8.02. This is almost obvious, and we shall only state on which type of relation the single reductions are based:

(i): 8.01.

(ii): 8.02 with $T = t_\sigma$.

(iii) and (iv): 8.01 in the form $(S_\sigma U_a S_\sigma^{-1}) \cdot t_\sigma (S_\sigma U'_a S_\sigma^{-1}) = t'_\sigma$ and 8.02 with $T = t'_\sigma$.

(v): 8.02 with $T = t'_\sigma$.

(vi): 8.02 with $T = \pi = S_\sigma U_a S_\sigma^{-1}$.

[If $\mathfrak{G}$ is a free product with one amalgamated subgroup, all the elements U_a, U_β belong to the one amalgamated subgroup $\mathfrak{U}$. Transforms of even length do not exist, i. e. a set ϕ_σ of generators of even length contains generators $f_{\sigma\tau}$ only, and relations involving these must be of the form 8.02 with $T = \pi$.]

In the general case, a relation 8.02 with $T = \pi$ and $l(f_{\sigma\tau})$ even entails for the central factors $C_{\sigma\tau}^{-1} U_a C_{\sigma\tau} = U_\beta (C_{\sigma\tau}$ in $\mathfrak{U})$ where $\mathfrak{U}_a$ and $\mathfrak{U}_\beta$ will, in general, be different groups. But this does not mean that U_a and U_β must both belong to the meet $\mathfrak{U}_{a\beta}$ of $\mathfrak{U}_a$ and $\mathfrak{U}_\beta$—as one might perhaps hope. One easily constructs examples to the contrary by taking $U_a = U_{\beta\gamma}$ in $\mathfrak{U}_{a\gamma} = \mathfrak{U}_a \cap \mathfrak{U}_\gamma$, and $U_\beta = U_{\beta\gamma}$ in $\mathfrak{U}_{\beta\gamma} = \mathfrak{U}_\beta \cap \mathfrak{U}_\gamma$, so that one may choose $C_{\sigma\tau}$ as an element in $\mathfrak{U}_\gamma$, which, in $\mathfrak{U}_\gamma$, transforms $U_{a\gamma}$ into $U_{\beta\gamma}$. And this is not the only possible type of example.

For the system of generators as it is at present, these relations cannot be simplified any further. However, in these relations there occur certain prime $\mathfrak{U}$-transforms, whose explicit representation by means of shorter generators we do not so far know. We only know particularly simple examples of prime $\mathfrak{U}$-transforms, viz. those generators of the form t_σ whose central factor, i. e. an element of a group $\mathfrak{G}_a$, or of $\mathfrak{U}$, is conjugate to an element of $\mathfrak{U}_a$ (or of any group $\mathfrak{U}_\beta$ in the latter case). It cannot be expected that every prime $\mathfrak{U}$-transform is of this form. In fact, not even those which occur in relations 8.0 need be generators. But in the following paragraphs we are going to show how, using the freedom left in the selection of the generators (cf. **3**), it is possible to choose them so that their defining relations 8.01 and 8.02 involve only prime $\mathfrak{U}$-transforms which are themselves generators of the form $t_{\sigma'}$.

To this end, we have to find out first which reduced words represent prime $\mathfrak{U}$-transforms. The next paragraph deals, more generally, with the corresponding question for prime words. The incidental results on prime words will be used throughout the later paragraphs.

9. From 7.6 and the definition of a prime word it is clear that the prime words of length l are those reduced words $P = \prod_1^n P_\nu$ whose longest generators are of length l, and which themselves are of the least possible length l.

We denote by $g_1, \cdots, g_m$ $(m \leq n)$ all the generators of length l in the order in which they appear in the reduced word $P \doteq \prod_1^n P_\nu$. If then P is to be of length l, the whole first and second half of all the generators $g_2, \cdots, g_{m-1}$ must cancel, and if l is odd, their central factors must amalgamate. By 4. 5 and 6. 2, this is possible only if all generators g_μ $(\mu = 1, \cdots, m)$ belong to the same set ϕ_σ, and $T(g_\mu) = S(g_{\mu+1})^{-1}$ for $\mu = 1, \cdots, m-1$. Besides, the product of shorter generators between g_μ and $g_{\mu+1}$ $(\mu = 1, \cdots, m-1)$ can be only the unit element, or a prime $\mathfrak{U}$-transform $\pi_\mu = T(g_\mu)^{-1} U_\beta T(g_\mu)$. As P is reduced, it follows that $m \leq 3$, and if either g_μ or $g_{\mu+1}$ (or both) are generators of the form t_σ, then certainly $\pi_\mu = 1$. Finally, the products of shorter generators preceding g_1, and following g_m, must not increase the length of the whole word, i. e. they must be prime words p_1 and p_2 respectively of length $< l$, so that $l(p_1 g_1) = l(g_m p_2) = l$. Now a distinction between the three possibilities for m gives immediately:

9. 0. *A reduced representation of a prime word p of length l is of one of the following types*:

9. 01. $p = p_1 g p_2$, *g a generator of length l*;

9. 02. $p^{\pm 1} = p_1 t_\sigma f_{\sigma\tau} p_2$;

9. 03. $p = p_1 f_{\sigma\tau_1}{}^{-1} \pi f_{\sigma\tau_2} p_2$, $\pi = S_\sigma U_a S_\sigma^{-1}$ *or* $= 1$;

9. 04. $p = p_1 f_{\sigma\tau_1}{}^{-1} t_\sigma f_{\sigma\tau_2} p_2$;

where in all cases p_1 and p_2 are prime words of length $< l$ with $l(p_1 g_1) = l(g_m p_2) = l$.

The first and second half of p depend only on p_1 and $S(g_1)$, and $T(g_m)$ and p_2, respectively. In particular, in order that there should exist a generator g' of the same length l as p such that $l(g'p) \leq l$ or $l(pg') \leq l$, the second or first half respectively of g' must cancel against p, and therefore against g_1 or g_m respectively (as shown in the proof of 7. 5). Hence by 6. 2, g' also belongs to ϕ_σ, and in the first case $p_1 = \pi_1 = S(g_1) U_{a_1} S(g_1)^{-1}$ (or $p_1 = 1$), and in the second case $p_2 = \pi_2 = T(g_m)^{-1} U_{a_2} T(g_m)$ (or $p_2 = 1$) are prime $\mathfrak{U}$-transforms.

Conversely, if $p_1 = \pi_1$, or $p_2 = \pi_2$ are of this form, then there exists a generator g' of length l such that $l(g'p) \leq l$ or $l(pg') \leq l$ respectively: one may simply take $g' = g_1^{-1}$ or $g' = g_m^{-1}$ respectively.

If g_1 or g_m is of the form t_σ, then, of course $\pi_1 = 1$, or $\pi_2 = 1$ respectively, since the representations of p are reduced. But if g_1 e. g. is of the form $f_{\sigma\tau_1}{}^{-1}$ and $p_1 = \pi_1 = T_{\sigma\tau_1}{}^{-1} U_{a_1} T_{\sigma\tau_1}$, we have

$$f_{\sigma\tau_1} \pi_1 f_{\sigma\tau_1}{}^{-1} = T_1 = S_\sigma (C_{\sigma\tau_1} U_{a_1} C_{\sigma\tau_1}{}^{-1}) S_\sigma{}^{-1},$$

hence T_1 is a $\mathfrak{U}$-transform, and either $T_1 = t_\sigma$, or $T_1 = \pi = S_\sigma U'_{a_1} S_\sigma{}^{-1}$, according as $l(T_1) = l$, or $l(T_1) < l$. Hence we can replace $\pi_1 f_{\sigma\tau_1}{}^{-1}$ by $f_{\sigma\tau_1}{}^{-1} T_1$, and then combine T_1 in the representation 9. 02, 9. 03, or 9. 04 with the next following factor which is of the same type as T_1. Correspondingly, if $g_m = f_{\sigma\tau_2}$, and $p_2 = \pi_2 = T_{\sigma\tau_2}{}^{-1} U_{a_2} T_{\sigma\tau_2}$, we replace $f_{\sigma\tau_2} \pi_2$ by $T_2 f_{\sigma\tau_2}$.

The representation of p thus obtained is in general no longer reduced, but has the advantage that it shows immediately whether there exists a generator of the same length l as the prime word p whose product with p is at most of length l. We call this the "principal form" of the prime word p.

9. 1. *A principal form of a prime word p of length l is of one of the following types:*

9. 11. $p = q_1 t_\sigma q_2$,

9. 12. $p^{\pm 1} = q_1 T f_{\sigma\tau} q_2$,

9. 13. $p = q_1 f_{\sigma\tau_1} T f_{\sigma\tau_2} q_2$,

where $T = t_\sigma$ or $T = \pi = S_\sigma U_a S_\sigma{}^{-1}$, or $T = 1$; and q_1 and q_2 are prime words of length $< l$ such that $l(q_1 g_1) = l(g_m q_2) = l$; q_1 and q_2 may be the unit, but $q_1 \neq S(g_1) U_{a_1} S(g_1)^{-1}$ and $q_2 \neq T(g_m)^{-1} U_{a_2} T(g_m)$.

9. 2. *There exists a generator g of the same length l as p such that $l(gp) = l$ or $l(pg) = l$ if, and only if, in the principal form of p, $q_1 = 1$ or $q_2 = 1$ respectively.*

We call q_1 and q_2 the first and second prime factor of p respectively.

In order to find the representations of prime $\mathfrak{U}$-transforms, we start with two lemmas which show that prime words behave very much like the actual generators:

9. 3. *If p and p' are two prime words of length l whose first prime factors in a principal form are one, and if P is a word in shorter generators such that $l(p^{-1} P p') \leq l$, then $l(p^{-1} p') \leq l$ and $P = \pi = S(p) U_a S(p)^{-1}$ is a prime $\mathfrak{U}$-transform of length $< l$.*

Proof. By 9. 2, there exist generators g and g' of length l so that

$l(g^{-1}p) \leq l$ and $l(g^{-1}p') \leq l$. Hence p and p' can be written so that $S(p) \equiv S(g)$ and $S(p') \equiv S(g')$; and if l is odd, also the central factors of p and g, and p' and g' belong to the same group.

In the products $g^{-1}P$ and Pg' less than half of g^{-1} and g' respectively are cancelled by P (by 5. 3); therefore the same holds for p^{-1} and p' in the products $p^{-1}P$ and Pp'. If $p^{-1}Pp'$ is to be of length $\leq l$, P must therefore be disposed of between p^{-1} and p', so that the whole first halves of p and p' can cancel against each other. But then we have with $l(p^{-1}Pp') \leq l$ also $l(g^{-1}Pg') \leq l$; hence by 6. 3 $l(g^{-1}g') \leq l$, i. e. $l(p^{-1}p') \leq l$, and

$$P = \pi = S(g)U_a S(g^{-1})^{-1} = S(p)U_a S(p)^{-1}.$$

An immediate consequence of 9. 3 is

9. 4. *If p and p' are any two prime words of length l such that $l(p^{-1}p') \leq l$, then there exist principal forms for p and p' whose first prime factors are equal, and whose first generators of length l, g_1 and g'_1 respectively, satisfy $l(g_1^{-1}g'_1) \leq l$. Correspondingly for the second prime factors and last generators of length l of two prime words satisfying $l(p'p^{-1}) \leq l$.*

Proof. Let $p = q_1\bar{p}$ and $p' = q'_1\bar{p}'$ be principal forms for p and p', i. e. $\bar{p}$ and $\bar{p}'$ the product of all the generators following the first prime factors q_1 and q'_1 respectively. Then $\bar{p}$ and $\bar{p}'$ are also prime words of length l, and $l(p^{-1}p') = l[\bar{p}^{-1}(q_1^{-1}q'_1)\bar{p}'] \leq l$, where $q_1^{-1}q'_1$ is a word in shorter generators, and $\bar{p}$ and $\bar{p}'$ are prime words whose first prime factors are one, i. e. they begin with generators of length l, g_1 and g'_1 respectively. By 9. 3 we have $l(\bar{p}^{-1}\bar{p}') \leq l$, hence also $l(g_1^{-1}g'_1) \leq l$. And:

$$q_1^{-1}q'_1 = \pi = S(g_1)U_a S(g_1)^{-1}, \text{ i. e. } q'_1 = q_1 \cdot \pi.$$

And in the principal form $p' = q'_1\bar{p}'$, π can be combined with the first generator g'_1 of $\bar{p}'$ as before. But then $p' = q_1(\pi\bar{p}')$ has the first prime factor q_1.

From 9. 1 together with 9. 4, the latter applied to the case that $p^{-1} = p'$, one obtains the principal forms of a prime 𝔘-transform of length l.

9. 5. *A principal form of a prime 𝔘-transform π of length l is of one of the following types:*

9. 51. $\pi = q_1 t_\sigma q_1^{-1}$ *where* $t_\sigma = S_\sigma(CU_aC^{-1})S_\sigma^{-1}$ *is a 𝔘-transform,*

9. 52. $\pi = q_1 f_{\sigma\tau}^{-1} T f_{\sigma\tau} q_1^{-1}$ *where* $T = t_\sigma = S_\sigma(CU_aC^{-1})S_\sigma^{-1}$ *or* $T = \pi_1 = S_\sigma U_a S_\sigma^{-1}$.

And from 9.2:

9.61. *There exists a generator g of length l such that $l(\pi g) \leq l$ or $l(g^{-1}\pi) \leq l$, if, and only if, $q_1 = 1$.*

9.62. *$\pi = t_\sigma$ is itself a generator of length l, if, and only if, there exists a generator $g = t'_\sigma$, or $g = f_{\sigma\tau}$, such that $l(\pi g) \leq l$.*

Proof. 9.61 is nothing but a restatement of 9.2 for prime U-transforms. The generator g in 9.61 can either be of the form $f_{\sigma\tau}{}^{-1}$, i.e. begin with $T_{\sigma\tau}{}^{-1}$ as first half, or of the form t'_σ or $f_{\sigma\tau}$, i.e. begin with S_σ as first half. In the first case, $\pi f_{\sigma\tau}{}^{-1} = t_\sigma f_{\sigma\tau}{}^{-1}$ is certainly of length $> l$, by 4.4. And conversely, if $g = t'_\sigma$ or $f_{\sigma\tau}$, and $\pi = t_\sigma$, then $l(\pi g) \leq l$.

Some further properties of prime words, again similar to those of generators, will be needed presently.

9.7. *If p is a prime word of length l, P a word in generators which are all shorter than l, then $l(p) \leq l(Pp) > l(P)$ and $l(p) \leq l(pP) > l(P)$.*

Proof. It is sufficient to prove one of the two inequalities, the first say.

By 7.4, a reduced representation of Pp contains as many generators of length l as that of p; hence, by 7.5: $l(Pp) \geq l = l(p)$.

To prove $l(Pp) > l(P)$, let the first prime factor of p in a principal form be q_1, i.e. $p = q_1 p_1$ where also p_1 is a prime word of length l such that, by 9.2, there exists a generator g of length l satisfying $l(g^{-1}p_1) \leq l$. In the product $Pp = Pq_1 p_1$, Pq_1 is a word in shorter generators, hence by 5.3 $l(Pq_1 \cdot g) > l(Pq_1)$, i.e. Pq_1 cancels less than $S(g)$, and cancels and amalgamates at most $S(g)$ in the product $Pq_1 \cdot g$. As we may assume $S(g) \equiv S(p_1)$, the same is true for p_1 in the product $Pq_1 \cdot p_1$, i.e. at least the last $\{\frac{1}{2}l\}$ factors of p_1 remain untouched and, if l is even, the last factor of $S(p_1)$ at most amalgamates. But $l(pp_1{}^{-1}) = l(q_1) < l$, and both, p and p_1, are of length l. Therefore, at least the last $\{\frac{1}{2}l\}$ factors of p and p_1 may be assumed to be identical and, if l is even, one more factor amalgamates. But then, in the product $Pq_1 \cdot p_1 = Pp$, at least the last $\{\frac{1}{2}l\}$ factors of p_1, i.e. of p, remain untouched and, if l is even, the last factor of $S(p)$ at most amalgamates; hence in fact $l(Pp) > l(P)$.

9.8. *If $g_1 = t_\sigma$ or $= f_{\sigma\tau}$ is a generator of the set ϕ_σ, and p a prime word shorter than g_1 such that $l(pg_1) = l(g_1)$, then also $l(pg_2) = l(g_2)$ for all other generators $g_2 = t'_\sigma$ or $= f_{\sigma\tau'}$ of ϕ_σ.*

Proof. As p is a word in shorter generators we have certainly

$l(pg_2) \geq l(g_2)$ by 5. 0. On the other hand we have $l(pg_1) = l(g_1)$ and $l(g_1^{-1}g_2) \leq l(g_2)$, hence by 2. 11: $l(pg_2) \leq l(g_2)$, and therefore $l(pg_2) = l(g_2)$.

9. 9. *If g is a generator of length l, p a prime word of length $l_1 < l$ whose principal form is $p = q_1 g_1 \cdots g_m q_2$, where $m \leq 3$ and $l(g_\mu) = l_1$, then with $l(pg) = l(g)$ also $l(q_2 g) = l(g_m q_2 g) = l(g)$, and with $l(gp) = l(g)$ also $l(gq_1) = l(gq_1 \cdot g_1) = l(g)$.*

Proof. It is sufficient to prove one of the two lines, the first say. If we write $p = \bar{p} q_2$, then $\bar{p}$ itself is also a prime word of length $l_1 < l$. Moreover $l(q_2) < l_1$; hence, by 2. 2, it follows from $l(\bar{p} q_2) = l(\bar{p})$ and $l(\bar{p} q_2 \cdot g) = l(g)$, that also $l(q_2 g) \leq l(g)$. But by 5. 0 $l(q_2 g) \geq l(g)$, hence $l(q_2 g) = l(g)$. $l(g_m q_2 g) = l$ follows in exactly the same way in the case $m > 1$, where p can be written in the form $p = \bar{p}(g_m q_2)$ with $\bar{p}$ itself a prime word of length l_1. But if $m = 1$, i. e. $p = q_1 g_1 q_2$, we proceed as follows: If we write $p = q_1 \bar{p}$, i. e. $\bar{p} = g_1 q_2$, then $l(\bar{p}) = l(p) < l(g)$ and $l(q_1) = l(q_1^{-1}) < l(p)$; hence $l(\bar{p} p^{-1}) < l(p^{-1})$ and $l(pg) = l(g)$, and therefore by 2. 11, applied to $\bar{p}^{-1}$, p^{-1}, g:

$$l(\bar{p} g) = l(g_1 q_2 \cdot g) \leq l(g).$$

But by 5. 0 $l(g_1 q_2 \cdot g) \geq l(g)$, hence $l(g_1 q_2 \cdot g) = l(g)$.

10. Up to now, we do not know how far the reduced representation of an element is unique. It seems best to deal with this question here in its natural context, although the results will not be needed until later.

We cannot expect the reduced representation $P = \prod_1^n P_\nu$ of an element of $\mathfrak{H}$ to be unique in the sense that two such representations 7. 1 of the same element are identical. For if, e. g., a factor P_ν is a generator of the form $P_\nu = t_\sigma$, and $P_{\nu+1}$ a word in shorter generators, hence $P_{\nu+1} \neq S_\sigma U_\alpha S_\sigma^{-1}$, then we can replace P_ν by $P_\nu \pi = t'_\sigma$, and $P_{\nu+1}$ by $\pi^{-1} P_{\nu+1}$, where π is any prime $\mathfrak{U}$-transform of the form $\pi = S_\sigma U_\alpha S_\sigma^{-1}$, and the resulting representation of P will still be reduced. We are not going to examine such possibilities in detail. All we shall need to know is, to what extent are the longest generators uniquely determined? We restrict ourselves to answering this question.

We begin by comparing two reduced representations of a prime word. The following proposition is analogous to the result 6. 2 for two generators of length l:

10. 0. *If p and p' are two prime words of length l, P a word in generators shorter than l, such that $l(p^{-1}Pp') < l$, then any two reduced representations*

of p and p' are of the same type (i. e. one of the four types 9. 01-9. 04), *the number of longest generators g_μ and g'_μ involved is the same for both, moreover it holds: if $g_\mu = f_{\sigma\tau}{}^{\epsilon}$, then $g'_\mu = g_\mu$ and if $g_\mu = t_\sigma = S_\sigma C_\sigma S_\sigma{}^{-1}$, then $g'_\mu = t'_\sigma = S_\sigma C'_\sigma S_\sigma{}^{-1}$, where C'_σ differs from C_σ at most by right- and left-hand factors in $\mathfrak{U}_a$, or of the form $C_{\sigma\tau} U_a C_{\sigma\tau}{}^{-1}$.*

Proof. We write the reduced representations of p and p' in the form $p = p_1 q p_2$ and $p' = p'_1 q' p'_2$, where p_1, p'_1 and p_2, p'_2 are the first and second prime factors of the reduced words p, p' respectively, i. e. all of them prime words of length $< l$, and q and q' are prime words of length l, both beginning and ending with a generator of length l. Then we have

$$p^{-1} P p' = (p_2{}^{-1} q^{-1})(p_1{}^{-1} P p'_1)(q' p'_2) = P' \text{ with } l(P') < l.$$

Hence also P' is a word in generators which are all shorter than l, and so are $p_1{}^{-1} P p'_1$ and $p_2 P' p'_2{}^{-1}$ in

$$q^{-1}(p_1{}^{-1} P p'_1) q' = p_2 P' p'_2{}^{-1}.$$

As q and q' are prime words of length l, whose first prime factors are one in a reduced representation (which then is also their principal form), we have by 9. 3

10. 01. $l(q^{-1}q') \leqq l$, and $p_1{}^{-1} P p'_1 = \pi = S(q) U_a S(q)^{-1}$ is of length $< l$. Hence, using $q(p_2 P' p'_2{}^{-1}) q'^{-1} = \pi$ in the same way, we have

10. 02. $l(qq'^{-1}) \leqq l$, and $p_2 P' p'_2{}^{-1} = \pi' = T(q)^{-1} U'_a T(q)$ is of length $< l$. By 10. 01 and 10. 02, $p^{-1} P p' = P'$ becomes

10. 03. $$q^{-1} \pi q' = \pi',$$

where q and q' contain the same generators of length l as p and p', so that it is sufficient to prove 10. 0 for these prime words whose prime factors are one, given that they satisfy 10. 03.

Because of 10. 01 and 10. 02, q and q' can be written with identical first and second halves,

$$q = S(q) C T(q) \text{ and } q' = S(q) C' T(q),$$

where C and C' belong to the same group; in fact, because of 10. 03,

10. 04. $$C^{-1} U_a C' = U'_a, \text{ i. e. } C' = U_a{}^{-1} C U'_a.$$

Moreover, $S(q) \equiv S(q')$ may be assumed to be identical with the first half of the first generator of q as well as q', and $T(q) \equiv T(q')$ identical with the second half of the last generator of q and q'.

Now we have to distinguish several cases:

1. If $S(q) = T(q)^{-1} = S_\sigma$, then both q and q' can only be of the type 9.01, i. e. $q = t_\sigma = S_\sigma C S_\sigma^{-1}$, $q' = t'_\sigma = S_\sigma C' S_\sigma^{-1}$, with $C' = U_a^{-1} C U'_a$.

2. If one of $S(q)^\epsilon$ and $T(q)^\epsilon$ equals S_σ, the other $T_{\sigma\tau}$, we may assume that $S(q) = S_\sigma$, $T(q) = T_{\sigma\tau}$. Then q and q' are either of the type 9.01 with $g = f_{\sigma\tau}$, or of the type 9.02, but with the same generator $f_{\sigma\tau}$, as there is no other generator with second half $T_{\sigma\tau}$ and a central factor in the same group as that of $f_{\sigma\tau}$. If q and q' were of different type, e. g. $q = f_{\sigma\tau}$, $q' = t_\sigma f_{\sigma\tau}$ with $t_\sigma = S_\sigma C_\sigma S_\sigma^{-1}$, then by 10.03: $q' = t_\sigma f_{\sigma\tau} = \pi^{-1} f_{\sigma\tau} \pi'$, i. e. q' would have been reducible by means of the reduction (iii). Hence they are of the same type, and either $q = q' = f_{\sigma\tau}$, or $q = t_\sigma f_{\sigma\tau}$, $q' = t'_\sigma f_{\sigma\tau}$, where because of 10.04:

$$C'_\sigma C_{\sigma\tau} = U_a^{-1} C U'_a = U_a^{-1} C_\sigma C_{\sigma\tau} U'_a, \text{ i. e. } C'_\sigma = U_a^{-1} C_\sigma (C_{\sigma\tau} U'_a C_{\sigma\tau}^{-1}),$$

as asserted.

3. Both $S(q)^{-1}$ and $T(q)$ are second halves of generators $f_{\sigma\tau}$, i. e. $S(q) = T_{\sigma\tau_1}^{-1}$, $T(q) = T_{\sigma\tau_2}$. Then q and q' are of one of the types 9.03 or 9.04, with the same generators $f_{\sigma\tau_1}$ and $f_{\sigma\tau_2}$. If q and q' were of different type, it follows as before from 10.03, that the one of type 9.04 would be reducible (this time by means of the reduction (iv)). Hence they are both of the same type; if this is the type 9.03, the assertions on the longest generators are obvious; if it is 9.04, we have $q = f_{\sigma\tau_1}^{-1} t_\sigma f_{\sigma\tau_2}$ and $q' = f_{\sigma\tau_1} t'_\sigma f_{\sigma\tau_2}$, hence by 10.04 $C' = C_{\sigma\tau_1}^{-1} C'_\sigma C_{\sigma\tau_2} = U_a^{-1} C U'_a = U_a^{-1} C_{\sigma\tau_1}^{-1} C_\sigma C_{\sigma\tau_2} U'_a$, i. e. $C'_\sigma = (C_{\sigma\tau_1} U_a^{-1} C_{\sigma\tau_1}^{-1}) C_\sigma (C_{\sigma\tau_2} U'_a C_{\sigma\tau_2}^{-1})$ for the central factors of t_σ and t'_σ. This completes the proof of 10.0.

10.0 includes in particular the case where $P = 1$, and p and p' are two reduced representations of the same prime word. Which shows, that in the case of prime words the longest generators involved in a reduced representation are, essentially, unique.

From this we want to deduce the corresponding result for any two reduced words which represent the same element of $\mathfrak{H}$:

10.1. *If* $P = \prod_1^n P_\nu$ *and* $Q = \prod_1^{n'} Q_\nu$ *are two reduced representations of the same element of* $\mathfrak{H}$, *and* $g_1, \cdots, g'_m$ *and* $g'_1, \cdots, g'_{m'}$ *respectively the generators of maximal length* l *in the order in which they occur in* P *and* Q, *then*

1. $m = m'$;

2. *if for some* μ $(1 \leq \mu \leq m)$; $g_\mu = f_{\sigma\tau}^{\epsilon}$, *then* $g'_\mu = g_\mu$;

3. *if for some* μ $(1 \leq \mu \leq m)$: $g_\mu = t_\sigma = S_\sigma C_\sigma S_\sigma^{-1}$, *then also* $g'_\mu = t'_\sigma = S_\sigma C'_\sigma S_\sigma^{-1}$, *where* C'_σ *differs from* C_σ *at most by left- and right-hand factors in* $\mathfrak{U}_\alpha$ *or of the form* $C_{\sigma\tau}^{-1} U_\alpha C_{\sigma\tau}$ (α *depends, in the usual manner, on the set* ϕ_σ).

Proof. The proof proceeds by induction, using 10. 0, i. e. the fact that we know the truth of these assertions for prime words. To that end we rewrite the representation 7. 1 of a reduced word $P = \prod_1^n P_\nu$ as a product of prime words of length l, and words in shorter generators, in the following way:

If the first factor P_1 is a word in shorter generators, hence P_2 a generator of length l, and if $l(P_1 P_2) > l$, we take P_1 as first factor of the modified representation; in all other cases we take the longest partial product $p_1 = \prod_1^{n_0} P_\nu$ which is of length l and therefore a prime word, as new first factor. If then P_{n_0+1} is a word in shorter generators, p_1 and P_{n_0+1} cancel less than half of each other, as otherwise also $\prod_1^{n_0+1} P_\nu$ would have been of length l. In that case, if $l(P_{n_0+1} P_{n_0+2}) > l$, we take P_{n_0+1} as next factor in the new representation, while in all other cases we take again the longest partial product $p_2 = \prod_{n_0+1}^{n_1} P_\nu$ which is of length l, and therefore a prime word, as next factor. So we go on, until finally the original representation 7. 1 of P appears written in the following form:

10. 11. $P = P_0 \prod_{\kappa=1}^{k} (p_\kappa P_\kappa)$, $k \geq 1$, *where the* p_κ $(\kappa = 1, \cdots, k)$ *are prime words of length* l, *and the* P_κ $(\kappa = 0, \cdots, k)$ *words in shorter generators or the unit. Any two successive factors* $\neq 1$ (*i. e.* p_κ, P_κ, *or* P_κ, $p_{\kappa+1}$, *if* $P_\kappa \neq 1$; *and* p_κ, $p_{\kappa+1}$, *if* $P_\kappa = 1$) *have the property that their product is longer than each factor* (*by* 9. 7 *and the rules for obtaining the form* 10. 11). *Moreover, if* $P_\kappa \neq 1$, *then the second prime factor of* p_κ *and the first prime factor of* $p_{\kappa+1}$ *are both one* (*since, as a factor of the representation* 7. 1, P_κ *was preceded and followed by a generator of length* l).

Now we write Q in the same way:

10. 12. $$Q = Q_0 \prod_{\kappa=1}^{k'} (q_\kappa Q_\kappa), \quad k' \geq 1.$$

Obviously, we may assume, that of all the different representations of this type for the same element of $\mathfrak{H}$, P (10. 11) is one with minimal k. Then $k' \geq k$.

As P and Q are representations of the same element of $\mathfrak{H}$, we have

10. 13. $$1 = P^{-1}Q = \prod_{\kappa=k}^{2} (P_\kappa^{-1} p_\kappa^{-1}) \cdot P_1^{-1} p_1^{-1} P_0^{-1} \cdot \ Q_0 q_1 Q_1 \prod_{\kappa=2}^{k'} (q_\kappa Q_\kappa).$$

Here, all generators of length l, hence all prime words on the right-hand side must cancel. Since at most the second half, i. e. the last $[\frac{1}{2}l]$ factors, of q_1 are cancelled from the right, at least the first $\{\frac{1}{2}l\}$ factors of q_1 must be cancelled from the left. We show that, thereby, also at least the last $\{\frac{1}{2}l\}$ factors of p_1^{-1} are cancelled, so that

10. 14. $$l(p_1^{-1} P_0^{-1} Q_0 q_1) < l.$$

This is seen as follows: Since at most the first $[\frac{1}{2}l]$ factors of p_1^{-1} are cancelled from the left, the assertion is obvious in the case that $P_0^{-1}Q_0 = 1$.

If $P_0^{-1}Q_0 \neq 1$, then, by 9. 7, at most half of it cancels on either side against less than $[\frac{1}{2}l]$ (if l is even), or against at most $[\frac{1}{2}l]$ (if l is odd) factors of p_1^{-1} and q_1. In order that the first $\{\frac{1}{2}l\}$ factors of q_1 should be cancelled from the left, $P_0^{-1}Q_0$ must, therefore, be eaten up between p_1^{-1} and q_1, i. e. its complete first and second halves must cancel on either side, and its central factor (if not in $\mathfrak{U}$) must cancel together with the two following factors on either side. Only then can the remaining factors of p_1^{-1} cancel the remaining of the first $\{\frac{1}{2}l\}$ factors of q_1. But, as p_1^{-1} and q_1 are of the same length l, an equal number of factors of p_1^{-1} and q_1 (and, after the above, at least $\{\frac{1}{2}l\}$) are cancelled in the product $p_1^{-1}P_0^{-1}Q_0q_1$, which proves 10. 14.

Therefore, $p_1^{-1}(P_0^{-1}Q_0)q_1 = \bar{P}$ with $l(\bar{P}) < l$; hence by 10. 0, the assertions 1.-3. of 10. 1 are true for the longest generators involved in the representations of p_1 and q_1 (both are reduced, by 7. 3); and these are the same as the first m_1 ($m_1 \leq 3$) generators of length l in the given reduced representations 7. 1 of P and Q respectively.

Besides, $\bar{P}$ is itself a word in shorter generators, and $P_0p_1 = Q_0q_1\bar{P}^{-1}$. After multiplication from the left with $(P_0p_1)^{-1}$, $P = Q$ becomes

$$P_1 \prod_{\kappa=2}^{k} (p_\kappa P_\kappa) = (\bar{P}Q_1) \prod_{\kappa=2}^{k'} (q_\kappa Q_\kappa).$$

Here the left-hand side is itself of the form 10. 11; and so is the right-hand side after, possibly, $\bar{P}Q_1$ has been combined with q_2 (viz. if the product PQ_1q_2 is of length l, and therefore itself a prime word). But the left-hand side involves only $k-1$ prime words of length l, so that an induction with respect to k completes the proof of 10. 1.

11. Now at last we can consider the problem of a more careful selection of the generators in order to simplify the relations 8. 0 as desired.

As in **3**, ϕ_0 is chosen as the set of all elements of length zero in $\mathfrak{H}$, i. e. it consists of all elements V of the meet $\mathfrak{B} = \mathfrak{H} \cap \mathfrak{U}$.

Assume that the sets $\phi_{\sigma'}$ of generators have been chosen for all ordinal numbers $\sigma' < \sigma$. Then $\mathfrak{K}_\sigma$ denotes again the group generated by all the generators of all the sets $\phi_{\sigma'}$ $(\sigma' < \sigma)$. If $\mathfrak{K}_\sigma$ is not yet the whole group $\mathfrak{H}$, the set ϕ_σ will be chosen as in **3**, but with an additional rule to be observed:

The first generator of ϕ_σ has to be an element of minimal length l of $\mathfrak{H} - \mathfrak{K}_\sigma$, and, if possible, a transform. Let $f_1, f_2, \cdots$ be all the elements of $\mathfrak{H} - \mathfrak{K}_\sigma$ which satisfy these first requirements (we choose this notation although they need not, of course, be denumerable), i. e. all transforms of length l, if transforms of length l exist in $\mathfrak{H} - \mathfrak{K}_\sigma$, otherwise all elements of length l in $\mathfrak{H} - \mathfrak{K}_\sigma$.

Let f be any one of these elements, g any generator of a set $\phi_{\sigma'}$ $(\sigma' < \sigma)$ already chosen. Then with f also gf belongs to $\mathfrak{H} - \mathfrak{K}_\sigma$, hence $l(gf) \geq l = l(f)$. We now consider all those earlier generators $g_{\sigma'} = t_{\sigma'}$ or $= f_{\sigma'\tau}$ (but not of the form $f_{\sigma'\tau}{}^{-1}$) with the property that

11. 00. $$l(g_{\sigma'}{}^{-1}f) = l.$$

All generators with this property are of length $< l$. For if there were one of the same length l as f, g' in $\phi_{\sigma'}$ say, then, by 3. 12, f would belong to $\mathfrak{K}_{\sigma'+1}$, and a fortiori to $\mathfrak{K}_\sigma$.

11. 01. *Put* $$\lambda_f = \operatorname{Max} l(g_{\sigma'})$$

where $g_{\sigma'}$ varies over all generators satisfying 11. 00. Then $\lambda_f < l$. If no generator $g_{\sigma'}$ satisfying 11. 00 exists, we put $\lambda_f = 0$. Finally, we put

11. 02. $\lambda_\sigma = \operatorname{Max} \lambda_f$, *where* f *varies over all elements* $f_1, f_2, \cdots$.

In the case that these elements $f_1, f_2, \cdots$ are transforms, we choose any one for which λ_f is maximal, i. e. $\lambda_f = \lambda_\sigma$, as first generator t_σ of ϕ_σ, and, as in **3**, we add all transforms t'_σ with $l(t'_\sigma t_\sigma) \leq l$ to ϕ_σ. I. e. $t_\sigma, t'_\sigma, \cdots$ are chosen so that:

11.10. *There exists an earlier generator* $g_{\sigma'}$ *of length* λ_σ *so that*

$$l(g_{\sigma'}^{-1}t_\sigma) = l = l(t_\sigma)\,;$$

11.11. *if also* $l(g_{\sigma''}^{-1}t_\sigma) = l$ *for some earlier generator* $g_{\sigma''} = t_{\sigma''}$ *or* $= f_{\sigma''\tau}$, *then* $l(g_{\sigma''}) \leq \lambda_\sigma$;

11.12. *If* t *is any other transform of length* l *in* $\mathfrak{H} - \mathfrak{K}_\sigma$, *and for some generator* $g_{\sigma^*} = t_{\sigma^*}$ *or* $f_{\sigma^*\tau}$, $l(g_{\sigma^*}^{-1}t) = l$, *then* $l(g_{\sigma^*}) \leqq \lambda_\sigma$.

If outside the group generated by $\mathfrak{K}_\sigma$ and the elements $t_\sigma, t'_\sigma, \cdots$ there are still elements f of length l such that $l(f^{-1}t_\sigma) \leq l$, we have to choose the next generator $f_{\sigma,1}$ of ϕ_σ from these elements f. Here again, we denote for every such f by $\lambda_{f,1}$ the length of a longest generator $g_{\sigma'} = t_{\sigma'}$ or $= f_{\sigma',\tau}$ (but $\neq f_{\sigma',\tau}^{-1}$) such that

$$l(fg_{\sigma'}) = l.$$

Then again $\lambda_{f,1} < l$ for all the elements f. Again we put

$$\lambda_{\sigma,1} = \operatorname*{Max}_{f} \lambda_{f,1},$$

and choose as next generator $f_{\sigma,1}$ of ϕ_σ an element f such that $\lambda_{f,1}$ is maximal, i. e. $\lambda_{f,1} = \lambda_{\sigma,1}$. Then the generator $f_{\sigma,1}$, besides satisfying 11.10-11.12 (because it has the same first half and a central factor out of the same group, as t_σ) has the property that:

11.20. *There exists an earlier generator* $g_{\sigma'}$[13] *of length* $\lambda_{\sigma,1}$ *so that*

$$l(f_{\sigma,1}g_{\sigma'}) = l = l(f_{\sigma,1})\,;$$

11.21. *if also* $l(f_{\sigma,1}g_{\sigma''}) = l$ *for some earlier generator* $g_{\sigma''} = t_{\sigma''}$ *or*

$$l(g_{\sigma''}) \leq \lambda_{\sigma,1}\,;$$

$= f_{\sigma''\tau}$, *then*

11.22. *if* f *is any other element of length* l *in* $\mathfrak{H} - \mathfrak{K}_\sigma$ *such that* $l(f^{-1}t_\sigma) \leq l$, *and if for some earlier generator* $g_{\sigma^*} = t_{\sigma^*}$ *or* $= f_{\sigma^*\tau}$, $l(fg_{\sigma^*}) = l$, *then* $l(g_{\sigma^*}) \leq \lambda_{\sigma,1}$.

So we go on: If all the generators t_σ, $f_{\sigma\tau'}$ for all $\tau' < \tau$ have been chosen, and there are still elements f of length l outside the group generated by $\mathfrak{K}_\sigma$ and all these generators t_σ and $f_{\sigma\tau'}$, such that $l(f^{-1}t_\sigma) \leq l$, we choose from these $f_{\sigma\tau}$ such that the length $\lambda_{\sigma,\tau}$ of a longest generator $g_{\sigma'}$ with $l(fg_{\sigma'}) = l$, is as great as possible. Again, 11.10-11.12 holds also for $f_{\sigma,\tau}$.

---

[13] Not normally the same as in 11.10, of course.

In the other case where $\mathfrak{H} - \mathfrak{K}_\sigma$ does not contain transforms of length l, we choose the generators correspondingly: first we see to it that λ_σ is maximal, i. e. $f_{\sigma,1}$, and all further generators of the set, will have to be chosen from those elements of length l of $\mathfrak{H} - \mathfrak{K}_\sigma$, for which the length λ_σ of a longest earlier generator $g_{\sigma'}$ with the property $l(g_{\sigma'}^{-1}f_{\sigma,1}) = l$, is maximal. Then we satisfy, step by step, the conditions that $\lambda_{\sigma\tau}$ (defined as above) is as great as compatible with the earlier choices. Again, λ_σ is the same for all generators $f_{\sigma\tau}$ of ϕ_σ, i. e. given by 11.10-11.12 (with $f_{\sigma\tau}$ in place of t_σ).

The gist of it is: In every step we see to it that as many factors as possible, both of the first, and of the inverse of the second half, of each generator are identical with the first half of some earlier generator. That this choice ensures the desired simplifications of the relations 8.0, we are going to show in the next paragraph.

We keep for these improved generators the old notation: ϕ_σ, etc.

12. All properties and results derived earlier for the generators of the system ϕ hold, of course, for our present generators. But the former result 5.2 can now be improved as follows:

12.0. *If g is a generator of length l, P a word in shorter generators such that $l(Pg) = l$ or $l(gP) = l$, then $P = p$ is a prime word of length $l_1 < l$ with the property that there exists a generator g_{σ_1} of length l_1, $g_{\sigma_1} = t_{\sigma_1}$, or $= f_{\sigma_1\tau}$, so that: $l(pg_{\sigma_1}) \leq l_1$ or $l(g_{\sigma_1}^{-1}p) \leq l_1$ respectively.*

Before we prove 12.0, let us consider the consequences: It follows immediately that the earlier result 6.2 can now be replaced by the following:

12.1. *If g_i and g_k are generators of equal length l, P a word in shorter generators, such that $l(g_iPg_k) \leq l$, then g_i and g_k belong to the same set ϕ_σ, $l(g_ig_k) \leq l$, and $P = \pi$ is a prime $\mathfrak{U}$-transform of length $l_1 < l$ which is itself a generator $t_{\sigma'}$ $(\sigma' < \sigma)$ of length l_1.*

Proof. By 6.2, $P = \pi$ is a prime $\mathfrak{U}$-transform of length $l_1 < l$, and $l(g_i\pi) = l(\pi g_k) = l$. Hence, using 12.0, there exists a generator $g_{\sigma'}$ of the form $g_{\sigma'} = t_{\sigma'}$ or $= f_{\sigma'\tau}$ of length l_1, such that $l(\pi g_{\sigma'}) \leq l_1$; as π is a transform, also $l(g_{\sigma'}^{-1}\pi) \leq l_1$. Therefore, by 9.62, $\pi = t_{\sigma'}$.

In particular, 6.3 is now improved to:

12.2. *If under the same assumptions as in 12.1, $l(g_iPg_k) < l$, then* *either* $g_i = t_\sigma$, $g_k = t'_\sigma$, *or* $g_i = f_{\sigma\tau}^{\epsilon}$, $g_k = f_{\sigma\tau}^{-\epsilon}$. *And then* $g_iPg_k = P'$, *where* $P = \pi = t_{\sigma'}$; $P' = \pi' = t_{\sigma''}$ $(\sigma' < \sigma,\ \sigma'' < \sigma)$.

And from this follows that the relations 8. 0 assume the following simplified form for our present generators:

12. 3. THEOREM. *Every relation between the generators of the system* ϕ (*consisting of the sets* ϕ_σ $(0 \leq \sigma < \sigma_0)$ *of generators* $t_\sigma^{(\rho)}$ *and* $f_{\sigma\tau}$ $(1 \leq \tau < \tau_\sigma)$) *follows from relations of the following two types*:

12. 31. $\prod_{\rho=1}^{r} t_\sigma^{(\rho)} = t_{\sigma'}$, *where* $t_{\sigma'} = S_\sigma U_\alpha S_\sigma^{-1}$ *is a generator of a set* $\phi_{\sigma'}$ *with* $0 \leq \sigma' < \sigma$;

12. 32. $f_{\sigma\tau}^{-1} t_{\sigma'} f_{\sigma\tau} = t_{\sigma''}$ *where* $t_{\sigma'} \neq 1$, $t_{\sigma''} \neq 1$ *are generators of the form* $t_{\sigma'} = S_\sigma(C_{\sigma\tau} U_\alpha C_{\sigma\tau}^{-1}) S_\sigma^{-1}$ *and* $t_{\sigma''} = T_{\sigma\tau}^{-1} U_\beta T_{\sigma\tau}$ *with* $0 \leq \sigma' \leq \sigma$; $0 \leq \sigma'' < \sigma$.

If the set ϕ_σ does not contain generators t_σ, then we have in 12. 32 necessarily $l(t_{\sigma'}) < l(f_{\sigma\tau})$, i. e. $\sigma' < \sigma$.

We denote this system of defining relations 12. 31 and 12. 32 for the generators of the system ϕ by $\mathfrak{R}$.

It now remains to prove 12. 0:

Proof of 12. 0. That under the assumptions of the theorem $P = p$ is a prime word of length $l_1 < l$, we know by 5. 2. All we have to prove is the additional property of the prime word p which can be expressed briefly as follows: If there is any generator g longer than the prime word p, such that the product pg is of the same length as this generator, then there is also a generator g_1 of the same length as the prime word and beginning with a half $S_{\sigma'}$ such that their product pg_1 is of the same length as this generator g_1. And correspondingly for the generator as left-hand factor.

12. 0 is trivially true for all generators g of the set ϕ_0 and all sets whose generators are of length one. Hence we may proceed by a transfinite induction, i. e. assume that 12. 0 is true for all products pg and gp where g belongs to a set $\phi_{\sigma'}$ with $\sigma' < \sigma$. We then have to prove it for all such products with a generator g of ϕ_σ.

Obviously it is sufficient to prove it for one of the two cases, e. g. $l(pg) = l$. The other case reduces to this one if one considers the inverse $p^{-1}g^{-1}$ which is also of length l.

First let us assume that g is of the form $g = t_\sigma$ or $g = f_{\sigma\tau}$, i. e. that g

begins with S_σ. In that case we may even assume that g is the first generator of ϕ_σ, since, by 9. 8, $l(pg_i) = l(g_i)$ for any generator $g_i = t_\sigma$ or $f_{\sigma\tau}$ of ϕ_σ implies the same for the first generator of ϕ_σ, and vice versa.

Now let $g_{\sigma'} = t_{\sigma'}$ or $= f_{\sigma'\tau}$ be a longest generator such that $l(g_{\sigma'}^{-1}g) = l$, then $l(g_{\sigma'}) = \lambda_\sigma$, where λ_σ is defined by 11. 10-11. 12. Also, of all the prime words shorter than g whose product with g as right-hand factor is of the same length l as g, let q be one of maximal length. Then we show:

12. 40. $$l(q) = \lambda_\sigma.$$

Proof. As $l(g_{\sigma'}^{-1}g) = l$, and $l(g_{\sigma'}^{-1}) = \lambda_\sigma$, and since $g_{\sigma'}^{-1}$ certainly is a prime word shorter than l, viz. of length λ_σ, we certainly have $l(q) \geqq \lambda_\sigma$. If q were longer than λ_σ i. e. $l(q) = \lambda > \lambda_\sigma$, we distinguish two cases:

Either q has the property that there exist a generator $g_{\sigma''}$ of the same length λ as q, and of the form $g_{\sigma''} = t_{\sigma''}$ or $f_{\sigma''\tau}$, such that $l(qg_{\sigma''}) \leq \lambda$. In that case $l(g_{\sigma''}^{-1}q^{-1}) \leqq l(q)$ and $l(qg) = l(g)$, hence by 2. 11: $l(g_{\sigma''}^{-1}g) \leqq l(g)$. But by 5. 0: $l(g_{\sigma''}^{-1}g) \geqq l(g)$, so that $l(g_{\sigma''}^{-1}g) = l(g)$. But $l(g_{\sigma''}) = \lambda > \lambda_\sigma$, which is a contradiction to 11. 11.

Or no such generator $g_{\sigma''}$ of the same length λ as q exists. Then, by 9. 1 and 9. 2, q is either of the form 9. 11 with $q_2 \neq 1$, or of one of the forms 9. 12 or 9. 13, where again $q_2 \neq 1$ unless q ends on a term of the form $f_{\sigma''\tau}q_2$, when the prime factor q_2 may, or may not, be one. Again we have to treat these cases separately:

If q is of the form $q = q_1 t_{\sigma''} q_2$ or $q = q_1 f_{\sigma''\tau}^{-1} q_2$, i. e. $q = q_1 g_{\sigma''}^{-1} q_2$ where $g_{\sigma''}$ begins with $S_{\sigma''}$, and $q_2 \neq 1$, we put $\bar{g} = q_2 g$. By 9. 9 $l(\bar{g}) = l(qg) = l$; and also

12. 41. $$l(g_{\sigma''}^{-1}\bar{g}) = l(\bar{g}) = l, \textit{ where } l(g_{\sigma''}) = \lambda > \lambda_\sigma.$$

And, should g be a transform, t_σ, then also $\bar{t} = q_2 g q_2^{-1}$ is a transform of the same length, and also

12. 42. $$l(g_{\sigma''}^{-1}\bar{t}) = l \textit{ with } l(g_{\sigma''}) = \lambda > \lambda_\sigma.$$

Besides, with g, also $\bar{g}$ or $\bar{t}$ respectively lie in $\mathfrak{H} - \mathfrak{K}_\sigma$, since they differ from g only by factors which are shorter than g, hence expressible by earlier generators. But then, 12. 41 or 12. 42 respectively is a contradiction to 11. 12. Hence this case is not possible.

If g is of one of the forms 9. 12 or 9. 13, and ends on a term $f_{\sigma''\tau} \cdot q_2$, then, again by 9. 9, with $l(qg) = l(g) = l$ also $l(f_{\sigma''\tau}q_2 \cdot g) = l$ and $l(q_2 g) = l$.

Hence, if we put $\bar{g} = f_{\sigma''\tau} q_2 \cdot g$, if g is no transform, or $\bar{t} = f_{\sigma''\tau} q_2 \cdot g \cdot q_2^{-1} f_{\sigma''\tau}^{-1}$, if $g = t_\sigma$ is a transform, then $l(f_{\sigma''\tau}^{-1} \cdot \bar{g}) = l = l(\bar{g})$, or $l(f_{\sigma''\tau}^{-1} \bar{t}) = l(\bar{t})$ respectively; besides $\bar{g}$, or $\bar{t}$, are elements of length l of $\mathfrak{H} - \mathfrak{K}_\sigma$, and we have the same contradiction to 11. 12 as before, as also $l(f_{\sigma''\tau}) = \lambda > \lambda_\sigma$. Hence $l(q) > \lambda_\sigma$ is impossible, and 12. 40 is proved: $l(q) = \lambda_\sigma = l(g_{\sigma'})$.

But then it follows from $l(g_{\sigma'}^{-1} g) = l(g)$ and $l(qg) = l(g)$, by 2. 12, that $l(qg_{\sigma'}) \leq l(g_{\sigma'}) = \lambda_\sigma$, which proves 12. 0 for prime words q of maximal length such that $l(qg) = l(g)$.

If now p is any other prime word such that also $l(pg) = l(g)$, then $l(p) < \lambda_\sigma = l(q)$, and again by 2. 12, it follows from $l(pg) = l = l(g)$ and $l(qg) = l = l(g)$ that $l(pq^{-1}) \leq l(q) = \lambda_\sigma$. As also $l(qg_{\sigma'}) = l(g_{\sigma'}) = \lambda_\sigma$, where $g_{\sigma'}$ has the same meaning as before, we have by 2. 11: $l(pg_{\sigma'}) \leq l(g_{\sigma'})$, hence, because of 5. 0: $l(pg_{\sigma'}) = l(g_{\sigma'})$. But $g_{\sigma'}$ is a generator of an earlier set $\phi_{\sigma'}$, for which we assumed 12. 0 to be true; i. e. there exists a generator $g_{\sigma''} = t_{\sigma''}$ or $= f_{\sigma''\tau}$ of the same length as p such that $l(pg_{\sigma''}) \leq l(p)$, q. e. d.

The proof in the case that g is of the form $g = f_{\sigma\tau}^{-1}$ is strictly analogous, using a generator $g_{\sigma'}$ of maximal length $\lambda_{\sigma\tau}$ with the property $l(f_{\sigma\tau} g_{\sigma'}) = l(g_{\sigma'}^{-1} g) = l$, as before. We shall not repeat it.

13. So far we have shown that any subgroup $\mathfrak{H}$ of the generalized free product $\mathfrak{G}$ can be generated by the system ϕ of generators (consisting of the sets ϕ_σ defined in **3**, **11**), whose defining relations are given by 12. 3. $\mathfrak{H}$ is a subgroup of $\mathfrak{U}$ if, and only if, ϕ consists of the one set ϕ_0 only. In that case the relations 12. 3 reduce to those of the form 12. 31, i. e. simply to the relations in the subgroup $\mathfrak{B}$ of $\mathfrak{U}$; and about those, they say nothing at all. In other words, our results give no information whatsoever on the subgroups of the 'smallest' generalized free product $\mathfrak{U}$.

If $\mathfrak{G}$ is larger than $\mathfrak{U}$, and the subgroup $\mathfrak{H}$ of $\mathfrak{G}$ not contained in $\mathfrak{U}$, it now remains to interpret the meaning of these relations.

13. 0. Theorem. *Every subgroup* $\mathfrak{H}$ *of* $\mathfrak{G}$, *not contained in* $\mathfrak{U}$, *is a generalized free product.*

13. 01. *Its factors are* (i) *subgroups* $\mathfrak{H}_\sigma$ *of conjugates of the groups* $\mathfrak{G}_\alpha$, (ii) *subgroups* $\mathfrak{H}^*_\sigma$ *of conjugates of* $\mathfrak{U}$, (iii) *groups which we denote by* $\mathfrak{F}_{\sigma\tau}$; *in* $\mathfrak{F}_{\sigma\tau}$ *the amalgamated subgroups generate a group which is self-conjugate with infinite cyclical factor group.*

13. 02. *All amalgamated subgroups are contained in groups generated by subgroups of conjugates of the groups* $\mathfrak{U}_\alpha$.

Proof. In conjunction with every single set ϕ_σ of the system ϕ of generators, we define the following subgroups of $\mathfrak{H}$:

1. We consider first the case that the generators of ϕ_σ are of odd length l. Let their central factors belong to $\mathfrak{G}_a - \mathfrak{U}_a$.

(a) If ϕ_σ contains generators $t_\sigma = S_\sigma C_a S_\sigma^{-1}, \cdots$, let $\mathfrak{H}_\sigma$ be the subgroup of $\mathfrak{H}$ which is generated by all these generators $t_\sigma, t'_\sigma, \cdots$. $\mathfrak{H}_\sigma$ contains all prime $\mathfrak{U}$-transforms π_σ of the form $\pi_\sigma = S_\sigma U_a S_\sigma^{-1}$. For with t_σ, also $\pi_\sigma t_\sigma = t'_\sigma$ is a generator of ϕ_σ. These elements π_σ form a subgroup of $\mathfrak{H}$ which we denote by $\mathfrak{T}_\sigma$.

Moreover, if $f_{\sigma\tau} = S_\sigma C_{\sigma\tau} T_{\sigma\tau}$ is any other generator of ϕ_σ (i. e. $C_{\sigma\tau}$ in $\mathfrak{G}_a - \mathfrak{U}_a$), and if $\mathfrak{H}_\sigma$ contains generators of the form

$$t_\sigma = s_\sigma{}^\tau = S_\sigma (C_{\sigma\tau} U_a C_{\sigma\tau}{}^{-1}) S_\sigma^{-1},$$

then let $\mathfrak{S}_\sigma{}^\tau$ be the subgroup of $\mathfrak{H}_\sigma$ which is generated by all these elements $s_\sigma{}^\tau$.

Finally, we denote by $\mathfrak{B}_\sigma$ the subgroup formed by all elements of $\mathfrak{H}_\sigma$ which are transforms of an element of $\mathfrak{U}_a$. All the groups $\mathfrak{S}_\sigma{}^\tau$, and $\mathfrak{T}_\sigma$, are then subgroups of $\mathfrak{B}_\sigma$.

(b) Next, whether ϕ_σ contains generators t_σ or not, we define for every generator $f_{\sigma\tau} = S_\sigma C_{\sigma\tau} T_{\sigma\tau}$ of ϕ_σ the following subgroups of $\mathfrak{H}$:

Let $\mathfrak{S}_{\sigma\tau}$ be the group of all elements $s_{\sigma\tau}$ of $\mathfrak{H}$ which can be written in the form $s_{\sigma\tau} = S_\sigma (C_{\sigma\tau} U_a C_{\sigma\tau}{}^{-1}) S_\sigma^{-1}$, i. e. $l(s_{\sigma\tau}) \leq l$ and if no generators t_σ exist in ϕ_σ, then certainly $l(s_{\sigma\tau}) < l$. Let $\mathfrak{T}_{\sigma\tau}$ be the group of all elements $t_{\sigma\tau}$ of $\mathfrak{H}$ which can be written in the form $t_{\sigma\tau} = T_{\sigma\tau}{}^{-1} U_a T_{\sigma\tau}$, i. e. $l(t_{\sigma\tau}) < l$.

Finally, we denote by $\mathfrak{B}_{\sigma\tau}$ the subgroup of $\mathfrak{H}$ which is generated by $\mathfrak{S}_{\sigma\tau}$ and $\mathfrak{T}_{\sigma\tau}$, and by $\mathfrak{F}_{\sigma\tau}$ the group generated by $\mathfrak{B}_{\sigma\tau}$ and $f_{\sigma\tau}$.

2. If the generators of ϕ_σ are of even length l, their central factors belong to $\mathfrak{U}$. I. e. in the case of a free product with one amalgamated subgroup $\mathfrak{U}$, there are certainly no transforms in ϕ_σ. Here the subgroups belonging to the generators $f_{\sigma\tau}$ of ϕ_σ are defined exactly as in 1.(b); and in this case $l(s_{\sigma\tau}) < l$ by necessity.

In the case of a generalized free product, the definitions are strictly analogous to 1.(a) and (b).

If ϕ_σ consists of transforms $t_\sigma = S_\sigma U_\sigma S_\sigma^{-1}$ (possibly) and generators $f_{\sigma\tau} = S_\sigma U_{\sigma\tau} T_{\sigma\tau}$, we denote by $\mathfrak{H}^*_\sigma$ the subgroup of $\mathfrak{H}$ generated by all t_σ, if these exist in ϕ_σ. It is a subgroup of a conjugate of $\mathfrak{U}$. Its subgroups $\mathfrak{T}_\sigma$, $\mathfrak{S}_\sigma{}^\tau$, and $\mathfrak{B}_\sigma$, as well as $\mathfrak{F}_{\sigma\tau}$ and its subgroups $\mathfrak{S}_{\sigma\tau}$, $\mathfrak{T}_{\sigma\tau}$, and $\mathfrak{B}_{\sigma\tau}$ are defined as in 1., always with $U_{\sigma\tau}$ in place of $C_{\sigma\tau}$, the significance of the suffixes of the groups $\mathfrak{U}_a$ which go into the definitions obtainable from 6.31 (this time from 6.31 (ii)).

Obviously, $\mathfrak{H}$ is generated by all the groups $\mathfrak{H}_\sigma$, $\mathfrak{H}^*{}_\sigma$ and $\mathfrak{F}_{\sigma\tau}$ $(\tau = \tau(\sigma))$. [In the case of one amalgamated subgroup $\mathfrak{U}$, the groups $\mathfrak{H}^*{}_\sigma$ do not occur.] Therefore the system ϕ' of all the elements t_σ and π_σ of all the groups $\mathfrak{H}_\sigma$ and $\mathfrak{H}^*{}_\sigma$, and all the elements $s_{\sigma\tau}$, $t_{\sigma\tau}$, and $f_{\sigma\tau}$ for all σ and $\tau = \tau(\sigma)$, is itself a system of generators for $\mathfrak{H}$. ϕ' includes the original generators t_σ and $f_{\sigma\tau}$ of ϕ. As regards the added generators π_σ, $s_{\sigma\tau}$, and $t_{\sigma\tau}$, it is obvious from their definition and 6.31, that the π_σ are exactly the prime $\mathfrak{U}$-transforms which occur in the relations 8.01, and the $s_{\sigma\tau}$ and $t_{\sigma\tau}$ are exactly the prime $\mathfrak{U}$-transforms which occur on the left- and right-hand side respectively of the relations 8.02. But in **12** we found that the system ϕ can be taken so that these prime $\mathfrak{U}$-transforms are themselves generators: the system $\mathfrak{R}$ of relations given in 12.3 enables us to express the new generators of the system ϕ' by means of generators of the original system ϕ:

13.11. $\pi_\sigma = t_{\sigma'}$ with $\sigma' < \sigma$,

13.12. $s_{\sigma\tau} = t_\sigma{}^\tau$, if $l(s_{\sigma\tau}) = l(f_{\sigma\tau})$,

$= t_{\sigma'}$ with $\sigma' < \sigma$, if $l(s_{\sigma\tau}) < l(f_{\sigma\tau})$,

13.13. $t_{\sigma\tau} = t_{\sigma''}$ with $\sigma'' < \sigma$.

By means of 13.11-13.13, the relations 12.31 and 12.32 of the system ϕ become:

13.21. $$\prod_{\rho=1}^{r} t_{\sigma}{}^{(\rho)} = \pi_\sigma,$$

13.22. $$f_{\sigma\tau}{}^{-1} s_{\sigma\tau} f_{\sigma\tau} = t_{\sigma\tau};$$

and these are now relations between the generators of one and the same group, viz. a group $\mathfrak{H}_\sigma$ or $\mathfrak{H}^*{}_\sigma$, or a group $\mathfrak{F}_{\sigma\tau}$.

We denote by $\mathfrak{R}'$ the system of all the relations 13.11-13.13 and 13.21-13.22 between the generators of the system ϕ'. ϕ' with the defining relations $\mathfrak{R}'$ is equivalent to ϕ with the defining relations $\mathfrak{R}$, i.e. $\mathfrak{H}$ can also be described by means of ϕ' and $\mathfrak{R}'$.

If we denote by $\mathfrak{H}_\beta, \mathfrak{H}_\gamma, \cdots$ unspecified members of the system of all the subgroups $\mathfrak{H}_\sigma$, $\mathfrak{H}^*{}_\sigma$, and $\mathfrak{F}_{\sigma\tau}$ of $\mathfrak{H}$, then the relations of $\mathfrak{R}'$ are of two different types: those of the form 13.21 or 13.22 are relations between generators of one and the same group $\mathfrak{H}_\beta$, while the relations of the form 13.11-13.13 are relations which identify an element of a group $\mathfrak{H}_\beta$ with an element of a group $\mathfrak{H}_\gamma$ $(\beta \neq \gamma)$. Now let h be any element of the meet of $\mathfrak{H}_\beta$ and $\mathfrak{H}_\gamma$. If, for the moment, we denote the generators of the system ϕ' which belong to $\mathfrak{H}_\beta$ and $\mathfrak{H}_\gamma$ respectively, by g_β and g_γ, we have $h = h(g_\beta) = h(g_\gamma)$. This is a relation between generators of the system ϕ', and follows, therefore, from the relations in $\mathfrak{R}'$. Hence we may add this relation, and in the same way any other relation which expresses that an element belongs to the meet of two groups $\mathfrak{H}_{\beta'}$ and $\mathfrak{H}_{\gamma'}$, to the system $\mathfrak{R}'$, and the enlarged system will still be a system of defining relations for the generators t_σ, π_σ, $s_{\sigma\tau}$, $t_{\sigma\tau}$, $f_{\sigma\tau}$ of $\mathfrak{H}$. But now this system of relations is of the kind described in I. 2. 0; hence $\mathfrak{H}$ is the generalized free product of all the groups $\mathfrak{H}_\sigma$, $\mathfrak{H}^*{}_\sigma$, and $\mathfrak{F}_{\sigma\tau}$.

Next we prove 13. 02 by showing:

13. 3. *The amalgamated subgroups are subgroups of* $\mathfrak{B}_\sigma$ *in the case of a factor* $\mathfrak{H}_\sigma$ *or* $\mathfrak{H}^*{}_\sigma$, *and of* $\mathfrak{B}_{\sigma\tau}$ *in the case of a factor* $\mathfrak{F}_{\sigma\tau}$.

By means of the relations 13. 11-13. 13, the reductions (i)-(vi) of **7** can now be based on the relations 13. 21 and 13. 22 instead of the relations 8. 01 and 8. 02. If then an element h of a factor $\mathfrak{H}_\beta$ of $\mathfrak{H}$ is somehow represented by means of the generators g_β of ϕ' which belong to $\mathfrak{H}_\beta$, then also the reduced representation will involve generators of this group $\mathfrak{H}_\beta$ only.

Now let h be an element of the meet of $\mathfrak{H}_\beta$ and $\mathfrak{H}_\gamma$. If one of these groups, $\mathfrak{H}_\beta$ say, is of the type $\mathfrak{H}_\sigma$ or $\mathfrak{H}^*{}_\sigma$, then h itself, as element of this group, is a generator $t_{\sigma'}$ $(\sigma' \leq \sigma)$, i. e. a reduced word which is unique by 10. 1. Hence, any reduction of the representation of h in $\mathfrak{H}_\gamma$ must lead to the same reduced representation, but for application of a relation of the type 13. 11-13. 12. This proves 13. 3 in this case.

If both $\mathfrak{H}_\beta$ and $\mathfrak{H}_\gamma$ are groups $\mathfrak{F}_{\sigma\tau}$ and $\mathfrak{F}_{\sigma'\tau'}$ respectively, then a representation of h in either group contains apart from generators which are $\mathfrak{U}$-transforms, and therefore of the form $t_{\sigma''}$, only the generators $f_{\sigma\tau}$ and $f_{\sigma'\tau'}$ respectively, and in each representation, this is a generator of maximal length. As $f_{\sigma\tau} \neq f_{\sigma'\tau'}$, it follows from 10. 1 that a reduced representation of h in either group cannot contain $f_{\sigma\tau}$ or $f_{\sigma'\tau'}$ any more, i. e. h is expressible by means of the generators $s_{\sigma\tau}$ and $t_{\sigma\tau}$, and $s_{\sigma'\tau'}$ and $t_{\sigma'\tau'}$ respectively. From which 13. 3 follows also in this case.

In the case of a factor $\mathfrak{F}_{\sigma\tau}$, the group generated by all its amalgamated subgroups is, in fact, exactly $\mathfrak{B}_{\sigma\tau}$. For, by 13. 12 and 13. 13, $\mathfrak{S}_{\sigma\tau}$ is amalgamated with the subgroup $\mathfrak{S}_\sigma{}^\tau$ of $\mathfrak{H}_\sigma$ or $\mathfrak{H}^*{}_\sigma$, if elements $s_{\sigma\tau}$ of length $l(f_{\sigma\tau})$ exist in $\mathfrak{S}_{\sigma\tau}$, otherwise with a subgroup of some earlier group $\mathfrak{H}_\sigma$ or $\mathfrak{H}^*{}_\sigma$; and $\mathfrak{T}_{\sigma\tau}$ is amalgamated with a subgroup of some group $\mathfrak{H}_{\sigma''}$ or $\mathfrak{H}^*{}_{\sigma''}$ $(\sigma'' < \sigma)$. Moreover, the relations 13. 21 show that $f_{\sigma\tau}$ transforms $\mathfrak{S}_{\sigma\tau}$

into $\mathfrak{T}_{\sigma\tau}$, i. e. $\mathfrak{V}_{\sigma\tau}$ is self-conjugate in $\mathfrak{F}_{\sigma\tau}$, and its factor group is the infinite cycle generated by $f_{\sigma\tau}$. This completes the proof of 13. 0.

In the case of a group $\mathfrak{H}_\sigma$ or $\mathfrak{H}^*_\sigma$, we know for certain only that the subgroups $\mathfrak{T}_\sigma$ and $\mathfrak{S}_\sigma{}^\tau$ are amalgamated subgroups: by 13. 11, $\mathfrak{T}_\sigma$ is amalgamated with a subgroup of some earlier factor $\mathfrak{H}_{\sigma'}$ of $\mathfrak{H}^*_{\sigma'}$, and each group $\mathfrak{S}_\sigma{}^\tau$ is amalgamated with the subgroup $\mathfrak{S}_{\sigma\tau}$ of the corresponding factor $\mathfrak{F}_{\sigma\tau}$. But these subgroups need not generate the whole of $\mathfrak{V}_\sigma$, which may contain other subgroups of conjugates of certain groups $\mathfrak{U}_\beta$, and these may, or may not, be amalgamated with subgroups of later groups $\mathfrak{H}_{\bar\sigma}$ or $\mathfrak{H}^*_{\bar\sigma}$ ($\bar\sigma > \sigma$).

All these results are the same in the case of a free product with one amalgamated subgroup, but for the fact, that the factors $\mathfrak{H}^*_\sigma$ which are subgroups of conjugates of $\mathfrak{U}$ in the general case, do not occur in this simpler case. In particular it follows:

13. 4. THEOREM.[14] *If* (i) $\mathfrak{G}$ *is a free product with one amalgamated subgroup* $\mathfrak{U}$, *and* (ii) $\mathfrak{U}$ *is self-conjugate in* $\mathfrak{G}$, *then any subgroup* $\mathfrak{H}$ *of* $\mathfrak{G}$ *is the free product of all the groups* $\mathfrak{H}_\sigma$ *and* $\mathfrak{F}_{\sigma\tau}$ *with the one amalgamated subgroup* $\mathfrak{V} = \mathfrak{H} \cap \mathfrak{U}$. (*For the meaning of* $\mathfrak{H}_\sigma$ *and* $\mathfrak{F}_{\sigma\tau}$ *cf.* 13. 01.)

Proof. $\mathfrak{U}$ is self-conjugate in $\mathfrak{G}$ if, and only if, $\mathfrak{U}$ is self-conjugate in every factor $\mathfrak{G}_\alpha$ of $\mathfrak{G}$.[15] Therefore all the groups $\mathfrak{T}_\sigma$, $\mathfrak{S}_\sigma{}^\tau$, $\mathfrak{V}_\sigma$, $\mathfrak{S}_{\sigma\tau}$, and $\mathfrak{T}_{\sigma\tau}$ are identical with the meet $\mathfrak{V}$ of $\mathfrak{H}$ and $\mathfrak{U}$, as any one of these groups consists of all elements of $\mathfrak{H}$ which are certain transforms of elements of $\mathfrak{U}$. And conversely, every element V of $\mathfrak{V}$ can be written as a transform $V = SUS^{-1}$ for any given S.

14. Finally we investigate the relation of the factors $\mathfrak{H}_\sigma$ and $\mathfrak{H}^*_\sigma$ to the given subgroup $\mathfrak{H}$. We need several preparatory lemmas.

14. 0. *If* K *is any element of* $\mathfrak{H}$ *which is a transform, then* $K = Q^{-1}t_\sigma Q$ *with* Q *in* $\mathfrak{H}$.

Proof. We use again the reduced representation of K written in the form 10. 11: if the longest generators are of length l, then

$$K = P_0 \prod_{1}^{k} (p_\kappa P_\kappa), \qquad k \geq 1,$$

where the p_κ ($\kappa = 1, \cdots, k$) are prime words of length l, and the P_κ

---

[14] Cf. Kurosch and Kalashnikov, [2], for a special case of 13. 4.

[15] Actually it can be seen without much trouble that the condition (i) of the theorem follows from (ii) provided that $\mathfrak{G} > \mathfrak{U}$.

($\kappa = 0, \cdots, k$) words in shorter generators, or $= 1$; any two successive factors $\neq 1$ have the property that their product is longer than each factor (Cf. **10**).

Moreover, we may assume that $P_0 = 1$; for otherwise we transform with P_0 which is an element of $\mathfrak{H}$; hence, if the lemma is true for the transform $P_0^{-1}KP_0$, then also for K.

If $k = 1$: $K = pP$. If here $P \neq 1$, then $l(pP) > l(p) = l$ and $l(pP) > l(P)$. As K is a transform, we have: $l(pP \cdot pP) < l(pP)$, i. e. at least the first half of the second factor pP is cancelled by the first factor pP. But because of $l(pP) > l(p)$, this means that at least the first $[\frac{1}{2}l] + 1$ factors of the second p are cancelled by the first factor pP, i. e. we also have

$$l(pPp) < l(pP), \text{ i.e. } l(Kp) < l(K).$$

But then either P alone cancels all there is cancelled of the right-hand factor p in this product, i. e. $l(Pp) < l(P)$ contrary to 9. 7; or else, P must be eaten up between the two factors p, so that the first p can then cancel the remaining factors of the second which are cancelled in the whole product pPp. But in pP more factors of P are untouched than are cancelled by p; hence, in Pp, more factors of P must cancel than remain unaffected, i. e. $l(Pp) < l(p)$, again in contradiction to 9. 7. Hence $P = 1$, $K = p$ is a prime word of length l, and 14. 0 follows from 9. 1, 9. 4, and 12. 1.

We assume 14. 0 to be true for all those transforms whose representation 10. 11 contains k' ($1 \leq k' < k$) prime words. Then let $K = \prod_1^k (p_\kappa P_\kappa)$ be a transform whose representation 10. 11 contains k prime words p_κ of length l.

Again we know, that in this representation at least the first $[\frac{1}{2}l] + 1$ factors of p_1 are untouched in the whole product $\prod_1^k (p_\kappa P_\kappa)$, and therefore are identical with as many factors of the first half of K. Since K is a transform, at least these $[\frac{1}{2}l] + 1$ factors of p_1 cancel in Kp_1. Again, as in the case $k = 1$, one deduces from this that $P_k = 1$. But then these first $[\frac{1}{2}l] + 1$ factors of p_1, which cancel in the product Kp_1, must do so against the last $[\frac{1}{2}l] + 1$ factors of p_k (because these also are untouched in the whole product $K = p_1P_1 \cdots P_{k-1}p_k$); hence we have $l(p_kp_1) < l$, so that p_kp_1 is expressible by generators which are all shorter than l, and

$$p_1^{-1}Kp_1 = P_1 \prod_2^{k-2} (p_\kappa P_\kappa) \cdot p_{k-1}P'_{k-1}$$

involves only $k - 2$ prime words of length l. By induction:

$$p_1^{-1}Kp_1 = Q_1^{-1}t_\sigma Q_1, \text{ i. e. } K = Q^{-1}t_\sigma Q,$$

where with Q_1 also $Q = Q_1 p_1^{-1}$ is an element of $\mathfrak{H}$.

14. 1. *Let* K *be again a transform in* $\mathfrak{H}$, *and* $Q, Q_1, Q_2, \cdots$ *all the elements of* $\mathfrak{H}$ *such that*

$$K = Q^{-1} t_\sigma Q = Q_1^{-1} t'_\sigma Q_1 = Q_2^{-1} t''_\sigma Q_2 = \cdots,$$

where $t_\sigma, t'_\sigma, t''_\sigma, \cdots$ *all belong to the same set* ϕ_σ. *If* Q *is one such element of minimal length, then* $l(t_\sigma Q) \geq l(Q)$.

Proof. If we had $l(t_\sigma Q) < l(Q)$; we put $\bar{Q} = t_\sigma Q$; then $\bar{Q}^{-1} t_\sigma \bar{Q} = Q^{-1} t_\sigma Q = K$; but $\bar{Q}$ is shorter than Q contrary to the assumption that Q is of minimal length.

If we put $t_\sigma = S_\sigma C_\sigma S_\sigma^{-1}$ and $K = S(K) C(K) T(K) = Q^{-1} t_\sigma Q$, where Q is again as short as possible, then Q^{-1} cancels at most S_σ, and cancels and amalgamates at most $S_\sigma C_\sigma$. Hence:

14. 11. $S(K) = Q^{-1} S_\sigma C$, *where* C *belongs to the same group as* $C(K)$;

14. 12. *Hence*: $C(K) = C^{-1} C_\sigma C$, *i. e.* C_σ *belongs to the same group as* $C(K)$.

From this we conclude:

14. 21. *If* K *and* K' *are two transforms of the same length in* $\mathfrak{H}$, *and* $l(KK') \leq l(K)$, *then* $K = Q^{-1} t_\sigma Q$, $K' = Q^{-1} t'_\sigma Q$, *where* t_σ *and* t'_σ *belong to the same set* ϕ_σ.

Proof. K and K' may be assumed to have identical first halves, and their central factors belong to the same group. By 14. 0, we have $K = Q^{-1} t_\sigma Q$ where we may assume Q to be of minimal length; hence, by 14. 11 and 14. 12:

$$S(K) = Q^{-1} S_\sigma C, \qquad C(K) = C^{-1} C_\sigma C,$$

and therefore $S(K') = Q^{-1} S_\sigma C$, $C(K') = C(K) C'$ (C' in the same group as $C(K)$). Hence: $K' = Q^{-1} S_\sigma (C_\sigma C C' C^{-1}) S_\sigma^{-1} Q = Q^{-1} t'_\sigma Q$.

The product $K_1 = KK'$, if shorter than K, is a prime $\mathfrak{U}$-transform. And because of 14. 21, $K_1 = Q^{-1}(t_\sigma t'_\sigma) Q$, which is a prime $\mathfrak{U}$-transform if, and only if, $t_\sigma t'_\sigma = \pi_\sigma = S_\sigma U_\alpha S_\sigma^{-1}$. If, on the other hand, K_1 and K are two transforms, $l(K_1) < l(K)$, such that $l(K_1 K) = l(K)$, then $K' = K_1 K$ is a transform of the same length as K, satisfying the assumption of 14. 21. Hence, 14. 21 applied to $K_1 K$ and K gives at once:

14. 22. *If* K *and* K_1 *are two transforms in* $\mathfrak{H}$, $l(K_1) \leq l(K)$, *such that* $l(K_1 K) = l(K)$, *then* $K_1 = Q^{-1} t'_{\sigma'} Q$, $K = Q^{-1} t_\sigma Q$, *where* $\sigma' = \sigma$ *if* $l(K_1)$

$= l(K)$; *and* $\sigma' < \sigma$, *but* $\pi_\sigma = t'_{\sigma'}$ *in the same group* $\mathfrak{H}_\sigma$, *or* $\mathfrak{H}^*_\sigma$, *as* t_σ, *if* $l(K_1) < l(K)$.

Finally we show that the set ϕ_σ of the generator t_σ in 14. 0 is, in general, uniquely determined by the transform K:

14. 3. *If* K *is a transform in* $\mathfrak{H}$, *but not a* $\mathfrak{U}$-*transform, and* $K = Q^{-1}t_\sigma Q = Q_1^{-1}t_{\sigma'}Q_1$, *then* $\sigma = \sigma'$.

Proof. If 14. 3 were not true, we may assume $\sigma' < \sigma$, i. e. $l(t_{\sigma'}) \leq l(t_\sigma)$. We have, with $QQ_1^{-1} = \bar{Q}$:

$$\bar{Q}^{-1}t_\sigma\bar{Q} = t_{\sigma'}.$$

Here $\bar{Q} \neq \pi = S_\sigma U_\alpha S_\sigma^{-1}$, since otherwise $\sigma = \sigma'$. We also assume $\bar{Q}$ to be a reduced word. $t_{\sigma'}$ is certainly reduced, hence unique by 10. 1. Reducing the left-hand side $\bar{Q}^{-1}t_\sigma\bar{Q}$ must, therefore, lead to the same result $t_{\sigma'}$.

If the longest generators in $\bar{Q}$ are shorter than t_σ, it follows immediately that $l(t_\sigma) = l(t_{\sigma'})$, and then $\sigma = \sigma'$ by 10. 1. If the longest generators in $\bar{Q}$ are at least of length $l(t_\sigma)$, all of them must drop out in the course of the reduction. But $\bar{Q}$, and its formal inverse $\bar{Q}^{-1}$, are reduced. Hence a reduction can arise only from t_σ in conjunction with the partial product Q_0g_1 of $\bar{Q}$ (where g_1 is the first generator of maximal length, and Q_0 a word in shorter generators or the unit element), i. e. $g_1^{-1}Q_0^{-1}t_\sigma Q_0g_1$ must be shorter than g_1, and therefore $Q_0^{-1}t_\sigma Q_0$ a prime $\mathfrak{U}$-transform, i. e. t_σ a $\mathfrak{U}$-transform, contrary to our assumption.

14. 0-14. 3 together give immediately:

14. 41. THEOREM. *If* $S^{-1}\mathfrak{G}_\alpha S$ *is a conjugate of one of the factors* $\mathfrak{G}_\alpha$, $\mathfrak{H}_S = \mathfrak{H} \cap (S^{-1}\mathfrak{G}_\alpha S)$ *its meet with* $\mathfrak{H}$, *then* $\mathfrak{H}_S$ *is conjugate in* $\mathfrak{H}$ *to a group* $\mathfrak{H}_\sigma$, $\mathfrak{H}_S = Q^{-1}\mathfrak{H}_\sigma Q$ *with* Q *in* $\mathfrak{H}$. $\mathfrak{H}_\sigma$ *is uniquely determined, unless* $\mathfrak{H}_S$, *and therefore* $\mathfrak{H}_\sigma$, *is contained in a conjugate of a group* $\mathfrak{U}_\alpha$.

14. 42. THEOREM. *If* $S^{-1}\mathfrak{U}S$ *is a conjugate of* $\mathfrak{U}$ *in* $\mathfrak{G}$, $\mathfrak{H}^*_S = \mathfrak{H} \cap (S^{-1}\mathfrak{U}S)$ *its meet with* $\mathfrak{H}$, *then* $\mathfrak{H}^*_S$ *is conjugate in* $\mathfrak{H}$ *to a group* $\mathfrak{H}^*_\sigma$, $\mathfrak{H}^*_S = Q^{-1}\mathfrak{H}^*_\sigma Q$ *with* Q *in* $\mathfrak{H}$. $\mathfrak{H}^*_\sigma$ *is uniquely determined, unless* $\mathfrak{H}^*_S$, *and therefore* $\mathfrak{H}^*_\sigma$, *is a subgroup of a conjugate of a group* $\mathfrak{U}_\alpha$.

In the case of a free product with one amalgamated subgroup, 14. 42 is, of course, irrelevant. These results again generalize Kurosch's results for ordinary free products.

15. If $\mathfrak{G}$ is the free product of the factors $\mathfrak{G}_\alpha$ with one amalgamated subgroup $\mathfrak{U}$, the essential points of our results are briefly these:

Every subgroup $\mathfrak{H}$ of $\mathfrak{G}$ is a generalized free product. The amalgamated subgroups are all contained in conjugates of $\mathfrak{U}$; the factors are of two different kinds: Either they are subgroups of conjugates of the factors $\mathfrak{G}_\alpha$ of $\mathfrak{G}$; these are, in general, uniquely determined, but for transformation in $\mathfrak{H}$. Or they are of the type $\mathfrak{F}_{\sigma\tau}$: the amalgamated subgroups of such a factor generate a self-conjugate subgroup of $\mathfrak{F}_{\sigma\tau}$ whose factor group is an infinite cycle.

These results, for the free product with one amalgamated subgroup, first led me to introduce the generalized free product. But the method used to investigate the simpler case yielded the corresponding results for the subgroups of a generalized free product $\mathfrak{G}$ with amalgamated subgroups $\mathfrak{U}_{\alpha\beta}$—as long as the subgroup is not contained in the smallest free product $\mathfrak{U}$. In this case, the factors of a subgroup $\mathfrak{H}$ are partly of the same types as before: factors $\mathfrak{F}_{\sigma\tau}$, and factors which are subgroups of conjugates of the $\mathfrak{G}_\alpha$. But in addition, we may have factors which are subgroups of conjugates of $\mathfrak{U}$. The amalgamated subgroups are subgroups of conjugates of the $\mathfrak{U}_\alpha$, i. e. the groups generated in each $\mathfrak{G}_\alpha$ by the $\mathfrak{U}_{\alpha\beta}$ (α fixed), which together generate $\mathfrak{U}$.

Now in $\mathfrak{G}$, $\mathfrak{U}$ together with any one group $\mathfrak{G}_\alpha$ generates a group which is the free product of $\mathfrak{U}$ and $\mathfrak{G}_\alpha$ with the one amalgamated subgroup $\mathfrak{U}_\alpha$. It is this property of $\mathfrak{G}$, rather than the full fact that $\mathfrak{G}$ is the generalized free product of the groups $\mathfrak{G}_\alpha$ with amalgamated $\mathfrak{U}_{\alpha\beta}$, which is reflected in our results on the subgroups $\mathfrak{H}$ of $\mathfrak{G}$. This is probably not due to the method only. The examples in Part I (8) gave us some indication that genuine difficulties are likely to arise in $\mathfrak{U}$. In this part (8), we came across one of the reasons: an element of $\mathfrak{U}$ may transform an element of one factor $\mathfrak{U}_\alpha$ into an element of another factor $\mathfrak{U}_\beta$, even if none of these elements belongs to the meet of $\mathfrak{U}_\alpha$ and $\mathfrak{U}_\beta$. It seems that the subgroups of $\mathfrak{U}$ cannot, in general, be represented as generalized free products derived from the corresponding representation of $\mathfrak{U}$. But we are not going to follow up this problem here.

UNIVERSITY COLLEGE, HULL.

REFERENCES.

[1] A. Kurosch, *Mathematische Annalen*, vol. 109 (1934), pp. 647-660.

[2] A. Kurosch and V. Kalaschnikov, *Comptes Rendus de l'Académie des Sciences, URSS*, vol. 1 (1935), pp. 285, 286 (quoted from: *Zentralblatt für Mathematik*, vol. 11 (1935), p. 151).

[3] B. H. Neumann, *Journal of the London Mathematical Society*, vol. 18 (1943), pp. 12-20.

SERIES TO SERIES TRANSFORMATIONS AND ANALYTIC CONTINUATION BY MATRIX METHODS.*

By P. VERMES.

---

1. Introduction. Several methods are known for the generalized summation by matrices of a series $u_0 + u_1 + \cdots$ with partial sums $s_k = u_0 + u_1 + \cdots + u_k$. The standard method is the *transformation of the sequence* s_k by a matrix $F \equiv (f_{nk})$ *into a convergent sequence* σ_n, so that $\sigma_n = \sum_k f_{nk} s_k$. The method is called *regular* if the existence of $\lim s_k$ implies the existence of $\lim \sigma_n$, and both limits are equal. We shall call the matrix of a regular sequence to sequence method a *T-matrix.* Sufficient and necessary conditions for a matrix to be a T-matrix are well known. (Dienes 389.) [1]

Another method is the *transformation of the series* $\sum u_k$ *into a convergent sequence* σ_n by a matrix $G \equiv (g_{nk})$, so that $\sigma_n = \sum_k g_{nk} u_k$. We shall call a matrix of a regular series to sequence method a *γ-matrix.* Sufficient and necessary conditions for G to be a γ-matrix are (Dienes 396):

$$\sum_k | g_{nk} - g_{nk+1} | \leq M \text{ for } n = 0, 1, 2, \cdots, \tag{1.1}$$

$$g_{nk} \to 1 \text{ as } n \to \infty \text{ for every fixed } k = 0, 1, 2, \cdots. \tag{1.2}$$

The third method is the *transformation of the series* $\sum u_k$ by the matrix $A \equiv (a_{nk})$ *into a convergent series* $\sum v_n$, so that $v_n = \sum_k a_{nk} u_k$. In section **2** we shall give sufficient and necessary conditions for the matrix A to be regular, in which case we shall call it an *α-matrix.* We also prove that the product of a γ-matrix and an α-matrix is a γ-matrix, and the product of two α-matrices is an α-matrix.

In section **3** we obtain the general expression for the matrix of the series to series transformation representing the continuation of a Taylor series. Such matrices form a one parameter family of upper-semi-matrices $a_{nk} = \binom{k}{n} \lambda^{k-n} (1 - \lambda)^n$. The corresponding sequence to sequence transformation matrices are $F = (1 - \lambda) A$. The product of matrices of the

---

* Received August 26, 1947.

[1] References are given at the end of this paper.

same family is commutative and associative and belongs to the family. Regularity, translative properties, equivalence of A- and F-methods depend on the parameter. There are also investigations on the field of efficiency for the series $\sum z^k$ and $\sum c_k z^k$ for simple and double transformations and for the inverses of the methods.

In section **4** we discuss the connection between the matrices of section **3** and those of the Euler summation methods $E(r)$ and $\mathcal{E}(r)$, elaborated recently by Agnew. The Euler matrices are transposed matrices of A and F. If A is an α-matrix its transpose is an Euler T-matrix, and if F is a T-matrix its transpose is an Euler α-matrix.

In section **5** we obtain the matrix A of the series to series transformation representing the continuation of the Laurent series. The result is a one parameter family of matrices $a_{nk} = \binom{k+n-1}{n} \lambda^k (1-\lambda)^n$, with corresponding matrices for sequence to sequence transformation $(1-\lambda)A/\lambda$. The properties of the matrices are investigated for various values of the parameter, and the field of efficiency for the series $\sum z^k$ and $\sum c_k z^k$. For certain values of λ the efficiency of the summation extends to values of z which are not in the principal star domain.

2. Series to series transformations, α-matrices.

The series $u_0 + u_1 + \cdots$ is summable by the matrix A to the sum s if

$$v_n = \sum_{k=0}^{\infty} a_{nk} u_k \text{ exists for } n = 0, 1, \cdots, \tag{2.1}$$

and

$$\sum_{n=0}^{\infty} v_n = s. \tag{2.2}$$

Writing $v_0 + v_1 + \cdots + v_n = \sigma_n$, (2.1) and (2.2) are equivalent to $\sigma_n \to s$ as $n \to \infty$. But

$$\sigma_n = \sum_{j=0}^{n} \sum_{k=0}^{\infty} a_{jk} u_k = \sum_{k=0}^{\infty} \Big(\sum_{j=0}^{n} a_{jk} \Big) u_k, \tag{2.3}$$

and writing

$$a_{0k} + a_{1k} + \cdots + a_{nk} = g_{nk}, \tag{2.4}$$

we have

$$\sigma_n = \sum_{k=0}^{\infty} g_{nk} u_k, \tag{2.5}$$

where the existence of (2.1) implies the existence of (2.5). Conversely it follows from (2.4) that

$$a_{0k} = g_{0k}, \qquad a_{nk} = g_{nk} - g_{n-1\,k} \qquad (n = 1, 2, \cdots), \tag{2.6}$$

and if σ_n, defined by (2.5), exists for $n = 0, 1, \cdots$,

$$v_n = \sum_{k=0}^{\infty} a_{nk} u_k = \sum_{k=0}^{\infty} (g_{nk} - g_{n-1\,k}) u_k = \sigma_n - \sigma_{n-1}, \tag{2.7}$$

so that the existence of (2.5) implies that of (2.1). Hence

2. I. *If the infinite matrices A and G are connected by the relation* (2.6) *or* (2.4), *then the series to series transformation by A and the series to sequence transformation by G are equivalent, i. e. the G-transforms are the partial sums of the series obtained by A-transformation.*

2. II. *The matrix $A \equiv (a_{nk})$ is an α-matrix if and only if G, given by $g_{nk} = a_{0k} + a_{1k} + \cdots + a_{nk}$, is a γ-matrix.*

It follows from (1.1), **2. II.**, (2.6) and (1.2) that

$$\sum_{k=0}^{\infty} | a_{nk} - a_{nk+1} | \leq M \text{ for } n = 0, 1, \cdots, \tag{2.8}$$

$$\sum_{n=0}^{\infty} a_{nk} = 1 \text{ for every } k, \text{ and hence } \lim_{n\to\infty} a_{nk} = 0. \tag{2.9}$$

Conditions (2.8) and (2.9) are satisfied by every α-matrix. They are not sufficient, as shown by the example

$$a_{nk} = 1 \text{ for } n \leq k; \ -1 \text{ for } k < n < 2k, \ 0 \text{ for } n \geq 2k. \tag{2.10}$$

If G is the lower semi-γ-matrix of ordinary convergence

$$G \equiv \begin{matrix} 1, 0, 0, \cdots \\ 1, 1, 0, \cdots \\ 1, 1, 1, \cdots \\ \cdot, \cdot, \cdot, \cdots \end{matrix}\ , \tag{2.11}$$

the unit matrix is the corresponding α-matrix.

2. III. *The product GA of a γ-matrix G and an α-matrix A exists and is a γ-matrix.*

Proof. It follows from (2.9) and from the regularity of G that

$$h_{nk} \equiv \sum_{j=0}^{\infty} g_{nj} a_{jk} \tag{2.12}$$

exists for every fixed n and k, and that

(2. 13) $$h_{nk} \to 1 \text{ as } n \to \infty \text{ for every fixed } k.$$

Also writing $a_{0k} + a_{1k} + \cdots + a_{nk} = b_{nk}$, B is a γ-matrix, and

$$h_{nk} = \sum_{j=0}^{\infty} g_{nj}(b_{jk} - b_{j-1\,k}) = \sum_{j=0}^{\infty} (g_{nj} - g_{nj+1}) b_{jk} + \lim_{j\to\infty} g_{nj} b_{jk},$$

where $b_{-1\,k} = 0$. It follows from (1. 1) that, as $j \to \infty$, g_{nj} tends to a finite limit, and from (1. 2) that $b_{jk} \to 1$, hence

(2. 14) $$\sum_{k=0}^{\infty} | h_{nk} - h_{nk+1} | \leq M_b M_g \text{ for every } n.$$

Thus, by (2. 12), $H = GA$ exists, and, by (2. 14) and (2. 13) it is a γ-matrix.

Note. The product AG may not exist, as can be seen from the example

$$A = \begin{matrix} 1, 1, 1, \cdots \\ 0, 0, 0, \cdots \\ 0, 0, 0, \cdots \\ \cdots\cdots \end{matrix} \qquad G = \begin{matrix} 1, 1, 1, \cdots \\ 1, 1, 1, \cdots \\ 1, 1, 1, \cdots \\ \cdots\cdots \end{matrix}$$

2. IV. *The product of two α-matrices exists and is an α-matrix.*

Proof. If A and B are α-matrices, and G, given by (2. 4), is a γ-matrix, then by **2. III.**, $H = GB$ is a γ-matrix, i. e. if $c_{nk} = h_{nk} - h_{n-1\,k}$ ($h_{-1\,k}$ being 0), then, by **2. II.**, C is an α-matrix. But

$$c_{nk} = \sum_{j=0}^{\infty} g_{nj} b_{jk} - \sum_{j=0}^{\infty} g_{n-1\,j} b_{jk} = \sum_{j=0}^{\infty} a_{nj} b_{jk},$$

hence $C = AB$ exists and is an α-matrix.

2. V. *A sufficient and necessary condition for the matrix product $H = GA$ to exist and be a γ-matrix for every γ-matrix G is that A should be an α-matrix.*

Proof. The sufficiency follows from **2. III.** To establish the necessity (for every γ-matrix G), we select the γ-matrix of example (2. 11), and then $h_{nk} = a_{0k} + a_{1k} + \cdots + a_{nk}$, so that H is a γ-matrix only if A is an α-matrix.

This result is parallel to a previous result (Vermes 1) which states that *a sufficient and necessary condition for the matrix product $H = AG$ to exist and be a γ-matrix for every γ-matrix G is that A should be a T-matrix.* The central position of the series to sequence method is demonstrated by the above results. This is also shown by the fact that if the matrices are triangular matrices, then for a given G the corresponding equivalent matrices F and A are

(2.15) $f_{nk} = g_{nk} - g_{nk+1}$ for sequence to sequence method,

(2.16) $a_{nk} = g_{nk} - g_{n-1\,k}$ for series to series method.

If G is a general matrix, the methods G and A are equivalent, but the methods G and F if and only if

(2.17) $$\lim_{n\to\infty} \left(\lim_{k\to\infty} g_{nk}s_k \right) = 0 \quad \text{(Vermes 4).}$$

2. VI. *If A_j are α-matrices $(j = 0, 1, \cdots, p)$ and $L = \beta_0 + \beta_1 + \cdots + \beta_p \neq 0$, then the matrix $H = (\sum \beta_j A_j)/L$ is an α-matrix.*

This follows immediately from a similar theorem on γ-matrices (Vermes 1, Theorem 1. II).

Thus if A and B are α-matrices and a and b any numbers such that $a + b \neq 0$, then there exists a unique α-matrix C such that $aA + bB = (a + b)C$. Hence an algebra of α-matrices can be formed in which the above *mean* represents sum, and matrix-product represents product. This algebra is incomplete, since the zero matrix is not an α-matrix.

If C is an α-matrix all of whose columns are equal, and A any α-matrix, then $CA = C$. If B is another α-matrix, then $CB = C$, hence $C(A - B) = 0$. Thus C has the left-hand zero property.

3. Continuation of the Taylor series as matrix transformation. A. Robinson has proved (Cooke 1) that if $f(z) = \sum a_n z^n$, with partial sums $s_n(z)$, has radius of convergence $r > 1$, and if for $0 < z' < 1$, $f(z) = \sum b_n (z - z')^n$, with partial sums $s'_n(z)$, then $s'_n(z'') = \sum f_{nk} s_k(z'')$, where $z'' > 1$, and $F \equiv (f_{nk})$ is a non-negative upper-semi-T-matrix.

In this section we develop a similar idea, considering in the first instance series to series transformations, obtaining simple expressions for classes of matrices depending on one parameter, and investigating some properties of these matrices. We also obtain simple expressions for the corresponding matrices for sequence to sequence transformations. The investigations are not restricted to regular matrices. Let

(3.1) $$f(z) = c_0 + c_1 z + c_2 z^2 + \cdots$$

be regular in a circle, and $z = \beta$ be a point inside this circle, so that

(3.2) $$f(z) = \sum_{n=0}^{\infty} (f^{(n)}(\beta)/n!)(z - \beta)^n, \text{ where } f^{(n)}(\beta)/n! = \sum_{k=n}^{\infty} \binom{k}{n} c_k \beta^{k-n},$$

for $n, k = 0, 1, 2, \cdots$ with the convention that

$$(3.3) \qquad \binom{k}{0} = 1 \text{ for every } k, \qquad \binom{k}{n} = 0 \text{ for } k < n.$$

If z' is in the circle of convergence of the series (3.2),

$$f(z') = \sum_{n=0}^{\infty} \sum_{k=n}^{\infty} \binom{k}{n} (\beta^{k-n}(z'-\beta)^n c_k z'^k)/(z'^{k-n} z'^n),$$

and writing $\beta/z' = \lambda$, we have

$$(3.4) \qquad f(z') = \sum_{n=0}^{\infty} \sum_{k=n}^{\infty} \binom{k}{n} \lambda^{k-n}(1-\lambda)^n c_k z'^k.$$

Thus the continuation (3.2) of the series (3.1) can be regarded as the generalized sum by series to series transformation of the series (3.1) by the upper semi-matrix A, where

$$(3.5) \qquad a_{nk} = \binom{k}{n} \lambda^{k-n}(1-\lambda)^n \qquad n, k = 0, 1, \cdots.$$

This matrix depends only on the ratio $\beta/z' = \lambda$, but not on the particular values of β or z', so that the same matrix represents continuations about other points to other values of z. It is independent of the coefficients c_k of the series. The parameter λ may take any complex value, and (3.5) defines a *family of matrices* which we shall denote by $A(\lambda)$. The series to series summation method by such a matrix will be called an $A(\lambda)$*-method.*

For any λ we have

$$(3.6) \qquad a_{0k} + a_{1k} + \cdots = 1 \text{ for a fixed } k.$$

This is equivalent to

$$(3.7) \qquad \lim_{n\to\infty} g_{nk} = 1, \ G \text{ being defined as in (2.4).}$$

A short calculation gives

$$(3.8) \qquad g_{nk} - g_{n\,k+1} = \binom{k}{n} \lambda^{k-n}(1-\lambda)^{n+1} \text{ for } k \geq n, \ = 0 \text{ for } k < n,$$

so that

$$(3.9) \qquad S_n \equiv \sum_{k=0}^{\infty} | g_{nk} - g_{n\,k+1} | = | 1-\lambda |^{n+1} \sum_{k=n}^{\infty} \binom{k}{n} | \lambda |^{k-n},$$

which converges if and only if $|\lambda| < 1$ or $\lambda = 1$.

Assuming (3.9) we have $S_n = |1-\lambda|^{n+1}/(1-|\lambda|)^{n+1}$ bounded if

(3.10) $$|1-\lambda| \leq 1-|\lambda|.$$

The last two conditions are simultaneously satisfied if and only if $0 \leq \lambda \leq 1$, and then the matrix G satisfies condition (1.1). Also (3.6) shows that (1.2) is satisfied. Hence we have proved:

3. I. *$A(\lambda)$ is an α-matrix if and only if $0 \leq \lambda \leq 1$.*

3. II. *$A(\lambda)$ is an α-matrix if and only if it is non-negative.*

The second result follows from **3. I** and the formula (3.5).

(3.11) *Definition.* If v_n denotes the A-transform of the series $u_0 + u_1 + \cdots$, and v'_n the A-transform of $0 + u_0 + u_1 + \cdots$, we say that A is *translative to the left* if the existence of $\sum v_n$ implies the existence of $\sum v'_n$ and the equality of both sums. If the reverse holds, we say that A is *translative to the right.* If A is translative to both the right and the left, we call it *translative.* Matrices possessing this latter property are called *regular* by Dienes, Cooke and author in their publications, named in the References. To denote 'translative to the left' they use the term *semi-regular.* Perron calls a translative method *permanent,* and Agnew uses the expression *A permits the adjunction or omission of elements.* The expressions adopted for this paper are taken over from Hill.

3. III. *$A(\lambda)$ is translative to the left.*

Proof. For $n \geq 1$, $v_n = \sum_{k=0}^{\infty} a_{nk}u_k$, $v'_n = \sum_{k=0}^{\infty} a_{n\,k+1}u_k$, so that for $n \geq 1$, $v'_n = \lambda v_n + (1-\lambda)v_{n-1}$, and $v'_0 = \lambda v_0$. Hence writing $v_0 + v_1 + \cdots + v_n = \sigma_n$, $v'_0 + v'_1 + \cdots + v'_n = \sigma'_n$,

(3.12) $$\sigma'_n = \lambda\sigma_n + (1-\lambda)\sigma_{n-1},$$

so that $\lim \sigma_n = s$ implies $\lim \sigma'_n = s$.

3. IV. *$A(\lambda)$ is translative to the right if $|1 - 1/\lambda| < 1$.*

Proof. From (3.12), writing $(\lambda - 1)/\lambda = b$, we obtain $b^{n-j}\sigma_j - b^{n-j+1}\sigma_{j-1} = b^{n-j}\sigma'_j/\lambda$ for $j = 0, 1, \cdots, n$, hence

(3.13) $$\sigma_n = (b^n\sigma'_0 + b^{n-1}\sigma'_1 + \cdots + \sigma'_n)/\lambda.$$

The right-hand side of (3.13) is the transform of the sequence σ'_n by the triangular matrix $f_{nk} = b^{n-k}/\lambda$ which is a T-matrix (Dienes 389) if $|b| < 1$, and then $\sigma'^n \to s$ implies $\sigma_n \to s$.

That the condition is not necessary can be seen from $\lambda = 0$, when A is the unit matrix, which is obviously translative.

The product of two matrices of the class $A(\lambda)$ always exists. Writing $1-\lambda = p$, and using the letters p, q, r in this sense, the matrix $A(p)$ shall denote

$$a_{nk} = \binom{k}{n} p^n (1-p)^{k-n}. \tag{3.14}$$

3. V. *The matrix product $A(p)A(q)$ is the matrix $A(pq)$.*

Proof. If $A(p)A(q) = B$,

$$b_{nk} = \sum_{j=n}^{k} \binom{j}{n} p^n (1-p)^{j-n} \binom{k}{j} q^j (1-q)^{k-j} (1-q)^{k-j} \text{ for } k \geq n, = 0 \text{ for } k < n.$$

But $\binom{k}{j}\binom{j}{n} = \binom{k}{n}\binom{k-n}{j-n}$ for $n \leq j \leq k$, thus

$$b_{nk} = \binom{k}{n} (pq)^n (1-pq)^{k-n},$$

and this proves the statement.

The matrices $A(p)$ form a class in which products are associative and commutative. If $p \neq 0$, $A(p)$ has a unique two-sided reciprocal $A(1/p)$ belonging to the same class. To $p = 1$ corresponds the unit matrix I, and to $p = 0$ the matrix

$$\Omega \equiv \begin{array}{l} 1, 1, 1, \cdots \\ 0, 0, 0, \cdots \\ 0, 0, 0, \cdots \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot \end{array}$$

We have $\Omega A(p) = A(p)\Omega = \Omega$.

Formula (3. 8) can be written in the form

$$g_{nk} - g_{n\,k+1} = (1-\lambda) a_{nk} = p \cdot a_{nk}, \tag{3.15}$$

which in view of (2. 15) states that the sequence to sequence summation matrix corresponding to $A(\lambda) = A(p)$ is given by

$$F(\lambda) = (1-\lambda) A(\lambda) = F(p) = p \cdot A(p). \tag{3.16}$$

It follows from (3. 15) and **3. V** that

3. VI. *The matrix product $F(p)F(q)$ is the matrix $F(pq)$.*

If $p = 1$, $F(p)$ is the unit matrix; if $p = 0$, $F(p)$ is the zero matrix.

For $p \neq 0$ $F(p)$ has the unique two-sided reciprocal $F(1/p)$. The matrices $F(p)$ form a class in which products are associative and commutative.

3. VII. *$F(\lambda)$ is a T-matrix if and only if $0 \leq \lambda < 1$.*

3. VIII. *$F(\lambda)$ is a T-matrix if and only if it is non-negative and different from the zero matrix.*

3. IX. *$F(\lambda)$ is translative to the right (left) if and only if $A(\lambda)$ is the same.*

The last three results are consequences of (3.16) and can be proved as the corresponding results for $A(\lambda)$.

(3.17) *Definition.* A summation method A is said to *include* the method B, if each series summable B is also summable A, the two generalized sums being equal.

The methods $A(\lambda)$ and $F(\lambda)$ are in general not equivalent. The relationship between them is revealed in the following:

3. X. *For $\lambda \neq 1$ $A(\lambda)$ includes $F(\lambda)$.*

Proof. The convergence of $\sum_k f_{nk}s_k = (1-\lambda)\sum_k a_{nk}s_k$ implies that as $k \to \infty$ $a_{nk}s_k \to 0$, and therefore, by (2.4), $g_{nk}s_k \to 0$ for every fixed n, so that (2.17) is satisfied.

3. XI. *A necessary condition for $F(\lambda)$ to include $A(\lambda)$ is that $|\lambda| < 1$.*

Proof. $A(\lambda)$ sums the series $1 + 0 + 0 + \cdots$ to 1. The $F(\lambda)$-transforms of the sequence of partial sums $1, 1, \cdots$ are $\sigma_n = (1-\lambda)^{n+1}\sum_{k=n}^{\infty}\binom{k}{n}\lambda^{k-n}$, and these converge to 1 only if $|\lambda| < 1$.

3. XII. *If the Taylor series $\sum c_k z^k$ has a non-zero radius of convergence, and if for $|\lambda| < 1$ $A(\lambda)$ sums the series in an open domain D, then $F(\lambda)$ sums the series to the same sum at all inner points of D.*

Proof. If z is in the circle of convergence, then (2.17) is obviously satisfied. If z is on or outside the circle, we can apply the following Lemma (Vermes 4, pp. 72-73):

LEMMA. *We suppose that $\sum c_k z^k$ has a finite nonzero radius of convergence, and that the G-method is efficient for the series in a domain D outside or on the circle of convergence. Then the transformation of the series by G and the transformation of the sequence s_k by F are equivalent if*

(3.18) $\lim_{k\to\infty} | g_{nk} |^{1/k}$ *exists and is finite for every n.*

In the present case g_{nk}, given by (2.4) and (3.5), satisfies condition (3.18), the limit being $| \lambda |$ for every n.

COROLLARY. *For any λ, $F(\lambda)$ includes $A(\lambda)$ on or outside the circle of convergence in the domain D.* This follows from the last part of the previous proof, where it was not required that $| \lambda | < 1$.

We shall now investigate the domain of efficiency of the method $A(\lambda)$ for the series $1 + z + z^2 + \cdots$, and then extend the result to partial star-domains.

3. XIII. *$A(\lambda)$ sums the series $1 + z + z^2 + \cdots$ if and only if z satisfies simultaneously the conditions*

$$| z | < 1/| \lambda |, \tag{3.19}$$

and

$$| 1 - \lambda | < | 1/z - \lambda |, \tag{3.20}$$

and the sum is the right value $1/(1 - z)$.

Proof. Applying the matrix $A(\lambda)$ to the series, we obtain

$$v_n(z) \equiv \sum_{k=0}^{\infty} a_{nk} z^k = \sum_{k=n}^{\infty} (1 - \lambda)^n \lambda^{-n} \binom{k}{n} (\lambda z)^k, \tag{3.21}$$

and excluding the trivial cases $\lambda = 0$ and $\lambda = 1$, the series converges if and only if (3.19) is satisfied, and then

$$\sum v_n(z) = \sum (z - \lambda z)^n / (1 - \lambda z)^{n+1}, \tag{3.22}$$

which requires (3.20).

We shall denote by $D(\lambda)$ the common part of the two circular domains of which (3.19) represents the inside of a circle with centre at the origin passing through the point $1/\lambda$, while the boundary of (3.20) is a circle passing through $z = 1$. The domain (3.20) is the inside or outside of the circle if the real part of λ is less or greater than 1/2, and it is a half-plane if $\mathcal{R}\lambda = 1/2$. For real $\lambda \leq 1/3$ (3.20) is inside (3.19), for $\lambda > 1/3$ the circles intersect. *$D(\lambda)$ contains the origin for every λ.*

The union of the domains $D(\lambda)$ for all values of λ is the union of all points to which the series can be continued, so that its boundary is the envelope of the circles passing through $z = 1$ and the centres of which are on the circle $|z| = 1$. The boundary, referred to polar coordinates with pole at $z = 1$, is given by $\rho = 4\sin^2\theta/2$.

The formulae (3.19), (3.20) can also be obtained by considering the continuation which lead to the matrix $A(\lambda)$ of (3.5). Writing $\beta = \lambda z'$, β is inside the circle of convergence, which is (3.19). The radius of convergence of the continued series is $|1 - \beta|$, and z' is in this circle if and only if $|z' - \beta| < |1 - \beta|$, which is (3.20). Hence $D(\lambda)$ is independent of the nature but dependent on the *location* of the singularity.

(3.23) *Definitions.* A simply connected domain in the z-plane is said to be a *star-domain,* if it contains the origin, and if every half-ray from the origin meets its boundary in at most one point. *$D(\lambda)$ is a star-domain,* as can be verified from (3.19) and (3.20).

If $f(z)$ is the function defined by $\sum c_k z^k$ and by its analytic continuations along half-rays from the origin, then the first singularity ζ reached on any half-ray is called a *vertex of the principal star-domain.* We denote by ζD the set of all points ζz for which z belongs to the set D. If D is a star-domain, having $z = 1$ as a boundary point, we shall call the common points of all domains ζD, formed from all vertices ζ of the principal star-domain, the *partial star-domain of $f(z)$ formed with D.* If D is bounded, and r' is the least, r'' the greatest distance of its boundary from the origin, and if R is the radius of convergence of $\sum c_k z^k$, then we can *modify* the definition of the *partial star-domain* in *restricting* the domains ζD to be considered to *all vertices ζ, for which $|\zeta| < Rr''/r'$.*

3. XIV. *$A(\lambda)$ sums the series $\sum c_k z^k$ in the partial star-domain formed with $D(\lambda)$, and the sum is $f(z)$.*

Proof. The summation represents the ordinary continuation. If $\phi(z)$ has a sole singularity at z', $\phi(z'v) \equiv \psi(v)$ has a sole singularity at $v = 1$, hence $A(\lambda)$ sums the Taylor series of $\phi(z)$ outside the region $z'D'(\lambda)$, D' being the complement of D with respect to the z-plane. Thus each singularity ζ requires the exclusion of a region $\zeta D'(\lambda)$. Since $D(\lambda)$ is a star-domain, the exclusion contains all regions belonging to singularities which may be beyond the vertex. The union of all $\zeta D'(\lambda)$ is the complement of the domain common to all $\zeta D(\lambda)$.

We shall now consider summation of the series $1 + z + z^2 + \cdots$ by two consecutive series to series transformations which correspond to a two-fold continuation. We first transform the series by $A(\lambda)$ into a series $\sum v_j(z)$ which may converge or diverge. Then we sum this series by the matrix $A(\lambda')$. We obtain $v_j(z)$ as in (3.22) if $|z| < 1/|\lambda|$. The series $\sum v_j(z)$ is then summable by $A(\lambda')$ if (3.19) and (3.20) are satisfied, replacing z by $(1-\lambda)/(1/z-\lambda)$ and λ by λ'. This gives the conditions:

$$(3.24)\qquad \text{(i)}\ |1/z| > |\lambda|, \qquad \text{(ii)}\ |1/z-\lambda| > |(1-\lambda)\lambda'|,$$
$$\text{(iii)}\ |1/z-\lambda-\lambda'+\lambda\lambda'| > |(1-\lambda)(1-\lambda')|.$$

We denote by $D[\lambda'(\lambda)]$ the common part of the three circles in (3.24), representing the domain of efficiency for the series $\sum z^k$ of the transformations by $A(\lambda)$ followed by $A(\lambda')$. The order, in which the transformations are carried out, is important: in general $D[\lambda'(\lambda)]$ is different from $D[\lambda(\lambda')]$, and both may differ from $D(\lambda\lambda')$, which is the domain of efficiency for the single summation by the product $A(\lambda)A(\lambda')$. The following *examples* may serve as illustration:

(3.25) $\lambda = 1/2$, $\lambda' = 1/3$. The repeated transformations both include $D(\lambda\lambda')$, and neither includes the other.

(3.26) $\lambda = \lambda' = -1/3$. All three domains are equal.

(3.27) $\lambda = 1/7$, $\lambda' = 1/3$, applied in this order reaches to the point $z = -7$, the greatest distance obtainable by two consecutive continuations.

(3.28) $\lambda = (1+i3^{\frac{1}{2}})/3$, $\lambda' = (4+i12^{\frac{1}{2}})/7$. $D[\lambda'(\lambda)]$ contains the real axis between $z=1$ and $z=3/2$, i.e. it extends partly outside the principal star-domain.

We now turn to the discussion of cases when the matrix $A(\lambda)$ has an inverse with respect to a Taylor series.

(3.29) *Definition.* If the matrix A transforms a series $\sum u_k$ into a series $\sum v_n$, and if a matrix B transforms the second series into the original series, then B is said to be the *inverse of A with respect to the series $\sum u_k$.*

If A has a two-sided reciprocal, the reciprocal is the inverse, provided that it applies to the transformed series. It will be convenient to use the notation of (3.14), when the reciprocal of $A(p)$ is $A(1/p)$.

3.XV. *$A(1/p)$ is for $p \neq 0$ the inverse of $A(p)$ with respect to $1+z+z^2+\cdots$ at the inner points of the segment*

$$(3.30)\qquad \text{(i)}\ |z| < 1/|1-p|, \qquad \text{(ii)}\ |z| < |1/(1-p)-z|.$$

Proof. This is a case of repeated transformations, but the last series need not converge. Hence (i) and (ii) of (3.24) is sufficient with $\lambda = 1-p$

and $\lambda' = 1 - 1/p$, and this gives (3.30). It can be verified that the second transformation then yields the original series.

The domain defined by (3.30), which we denote by $S(p)$, is the segment of a circle with centre at the origin and passing through $z = 1/(1-p)$ cut off by a line that bisects perpendicularly the line joining the origin to $1/(1-p)$. For $p < 1$, $S(p)$ contains the singular point $z = 1$, and for suitable p may extend to any point of the z-plane. *The union of domains* $S(p)$ *for all values of* p *is the whole* z*-plane.*

3. XVI. $A(1/p)$ *is for* $p \neq 0$ *the inverse of* $A(p)$ *with respect to the series* $\sum c_k z^k$ *in the partial star-domain formed with* $S(p)$.

This follows from **3. XV** in the same way as **3. XIV** from **3. XIII.** Although the boundary of $S(p)$ does not pass through $z = 1$, $S(p)$ is a star-domain, and the definition can be extended to this case.

COROLLARY 1. *If* R *is the radius of convergence of the series* $\sum c_k z^k$, *the partial star-domain contains the circle* $|z| < R/2\,|1-p|$.

This follows from the remark at the end of (3.23), since for $S(p)$ $r' = 1/2\,|1-p|$. If $R \neq 0$, we can make the partial star-domain arbitrarily large. For $z = 1$ we then obtain:

COROLLARY 2. *If* $\overline{\lim}\,|u_k|^{1/k} = 1/R$ *is finite,* $A(p)$ *has an inverse* $A(1/p)$ *with respect to* $\sum u_k$ *for all* p *satisfying* $|1-p| < R/2$.

COROLLARY 3. *The theorem holds for* $F(p)$ *and sequence* $c_k z^k$.

4. The transpose of the Taylor series continuation matrix. The transpose A' of a matrix A is defined by $a'_{nk} = a_{kn}$. The transpose of $A(p)$ is the matrix

$$a'_{nk} = \binom{n}{k} p^k (1-p)^{n-k} \tag{4.1}$$

with the convention (3.3), a triangular matrix, which is the matrix $E(p)$ of the Euler sequence to sequence transformation (Agnew, formula 1.2).

4. I. *The transpose* $E(p)$ *of the matrix* $A(p)$, $p \neq 0$, *is a* T*-matrix if and only if* $A(p)$ *is an* α*-matrix.*

This follows from the well-known result that $E(p)$ is regular if and only if $0 < p \leq 1$ (Agnew, page 314), and from **3. I.** The transpose of the matrix $F(p)$ is the matrix

$$f'_{nk} = \binom{n}{k} p^{k+1} (1-p)^{n-k} = p a'_{nk}, \tag{4.2}$$

a triangular matrix, which is the matrix $\mathcal{E}(p)$ of the Euler series to series transformation (Agnew, formula 7. 2).

4. II. *The transpose* $\mathcal{E}(p)$ *of the matrix* $F(p)$ *is an* α*-matrix if and only if* $F(p)$ *is a* T*-matrix.*

Proof. Considering the series to sequence transformation matrix G corresponding to $\mathcal{E}(p)$ given as in (2. 4) we obtain

$$\lim_{n\to\infty} g_{nk} = \sum_{j=k}^{\infty} \binom{j}{k} p^{k+1}(1-p)^{j-k}, \tag{4. 3}$$

and, as calculated in Agnew's paper (7. 61 and 7. 62),

$$\begin{aligned} \sum_{k=0}^{\infty} \left| g_{nk} - g_{n\,k+1} \right| &= \sum_{k=0}^{n} \binom{n+1}{k+1} (1-p)^{n-k} p^{k+1} \Big| \\ &= (|1-p| + |p|)^{n+1} - |1-p|^{n+1}, \end{aligned} \tag{4. 4}$$

so that (1. 1) and (1. 2) are satisfied if and only if $0 < p \leq 1$, which is the condition for $F(p)$ to be a T-matrix.

From (4. 4) follows that $\mathcal{E}(p)$ is the matrix corresponding not to $E(p)$ but to a matrix $(a'_{n+1\,k+1})$, obtained from $E(p)$ by removing its first row and first column. The two methods $E(p)$ and $\mathcal{E}(p)$ are not equivalent, as can be seen from the *example*:

Applying $E(p)$ to the sequence $1, 1, \cdots$ we obtain $\sigma_n = 1$ for every n, i. e. the $E(p)$-sum of the series $1 + 0 + 0 + \cdots$ is 1 for every p. But the $\mathcal{E}(p)$-transform of this series converges only if $|1-p| < 1$. This follows also from Theorem 7. 8. of Agnew's paper, which gives a complete discussion of the two methods.

There is one point we wish to pursue. Agnew (Theorem 8. 2) proves that the series $\sum c_k z^k$ is summable by $E(p)$ in the partial star-domain formed with $D(p)$ (cf. (3. 23) of this paper), where

(4. 6) $D(p)$ *is given by* $|z - (1 - 1/p)| < |1/p|$ (Agnew 8. 11), *provided that* $|1-p| < 1$.

We shall investigate the case when $|1-p| > 1$, mainly for $\mathcal{E}(p)$-summability. It has been shown by Agnew (pp. 327-328) that the series $\sum z^k$ is summable $\mathcal{E}(p)$ for $p \neq 0$ in the circle (4. 6). Our aim is to state sufficient conditions under which the series $\sum c_k z^k$ is summable $\mathcal{E}(p)$ when $|1-p| > 1$. In the case treated by Agnew the origin is inside the circle (4. 6), so that $D(p)$ is a star-domain, while in the present case the origin is outside.

4. III. *If the function $f(z)$, represented by the series $\sum_0^\infty c_k z^k$ and its analytic continuations, has a single finite singularity at $z = \zeta \neq 0$, then it is summable $\mathcal{E}(p)$ for $|1-p| > 1$ in the circle $\zeta D(p)$, provided that $f(1/z)$ tends to zero uniformly as $z \to 0$.*

Proof. If z is in $\zeta D(p)$, then z/ζ is in $D(p)$, hence (4. 6) gives

$$(4.7) \qquad |1-p+pz/\zeta| < 1 \text{ for all } z \text{ in } \zeta D(p).$$

The functions of the complex variable w

$$(4.8) \qquad \phi(w) \equiv f(w)/(w-z), \qquad \psi(w) \equiv pw^{-1}f(w)\sum_{n=0}^{\infty}(1-p+pz/w)^n$$

are equal for all values of w satisfying

$$(4.9) \qquad |1-p+pz/w| < 1,$$

hence their residues are equal.

Denoting the residue of $\phi(w)$ at $w = \alpha$ by $R(\phi, \alpha)$, (4. 8) gives

$$(4.10) \qquad R(\phi, \zeta) + R(\phi, z) = R(\psi, \zeta) + R(\psi, 0).$$

Now, $R(\phi, z) = f(z)$, and at $w = \zeta$ (4. 9) is satisfied, hence $R(\phi, \zeta) = R(\psi, \zeta)$. Also substituting for $f(w)$ the series $\sum c_k w^k$, and for $(1-p+pz/w)^n$ its binomial expansion, (4. 10) becomes

$$(4.11) \qquad \sum_{n=0}^{\infty}\sum_{k=0}^{n}\binom{n}{k}p^{k+1}(1-p)^{n-k}c_k z^k = f(z),$$

whenever the series converges. Considering the functions of the complex variable t

$$(4.12) \qquad V_n(t) \equiv pt^{-1}f(1/t)(1-p+pzt)^n \qquad n = 0, 1, \cdots,$$

for $|t| \geq r > 1/|\zeta|$ we have $t^{-1}f(1/t) = c_0/t + c_1/t^2 + \cdots$, hence integrating term by term we obtain

$$(4.13) \qquad (1/2\pi i)\int_{|t|=r} V_n(t)\,dt = \binom{n}{k}p^{k+1}(1-p)^{n-k}c_k z^k \equiv v_n(z).$$

The only singularities of the integrand in and on the path of integration are at $t = 1/\zeta$ and $t = 0$, hence as in (4. 10)

$$(4.14) \qquad v_n(z) = R(V_n, 1/\zeta) + R(V_n, 0).$$

For a fixed z satisfying (4. 7), and for $|1-p+pz/\zeta| < \theta < 1$

$$(4.15) \qquad |R(V_n, 1/\zeta)| \leq K\theta^n,$$

K being independent of n. Also, for a sufficiently small positive ρ and $|t| \leq \rho$ $|1 - p + pzt|^n \leqq K'$, and since $f(1/t) \to 0$ uniformly as $t \to 0$,

$$R(V_n, 0) = 0. \tag{4.16}$$

From (4.14), (4.15) and (4.16) follows the convergence of $\sum v_n(z)$, and (4.11) is established.

COROLLARY 1. *If* $f(z)$ *has several singularities* ζ $(\neq 0)$, *it is summable* $\mathcal{E}(p)$ *in the domain common to all* $\zeta D(p)$.

COROLLARY 2. *Under the conditions of the theorem or of Corollary* 1 *the series is summable* $E(p)$. This follows from Agnew, Theorem 7.8, which states that $E(p)$ includes $\mathcal{E}(p)$.

Examples.

(4.17) If $f(z) \equiv (1 - z)^{-2}$ given by $1 + 2z + 3z^2 + \cdots$, we have $v_n(z) = p(1 - p + pz)^n + np^2z(1 - p + pz)^{n-1}$, hence in $D(p)$ $(p \neq 0)$ $\sum v_n(z) = (1 - z)^{-1} + z(1 - z)^{-2} = (1 - z)^{-2}$. Here $f(1/z) \to 0$ uniformly.

(4.18) If $f(z) \equiv z^2(1 - z)^{-2}$ given by the series $0 + 0 + z^2 + 2z^3 + \cdots$, $v_n(z) = np^2z(1 - p + pz)^{n-1} - p(1 - p + pz)^n + p(1 - p)^n$, and $\sum v_n(z)$ converges in $D(p)$ if and only if $|1 - p| < 1$. Here $f(1/z) \to 1$.

(4.19) If $f(z) \equiv -z^{-1} \log(1 - z)$ given by $1 + z/2 + z^2/3 + \cdots$, $v_n(z) = [(1 - p + pz)^{n+1} - (1 - p)^{n+1}]/(n + 1)z$, and $\sum v_n(z)$ converges in $D(p)$ if and only if $|1 - p| < 1$. Here $f(1/z) \to 0$ as $z \to 0$, but not *uniformly*, considering all branches of the multivalued function.

5. Continuation of the Laurent series as a matrix transformation.[2] In this section we first consider the series to series transformation which represents the continuation of the Laurent series by a Taylor series. We obtain families of matrices similar to those in section **3** with corresponding expressions for sequence to sequence transformation matrices.

(5.1) Let $f(z) = \sum_{k=0}^{\infty} c_k(\beta - z)^{-k}$ represent a function regular at all finite points of the z-plane except at $z = \beta \neq 0$. Expressed as a Taylor series $\sum C_n z^n$ in the circle $|z| < \beta$, the coefficients C_n, obtained from the expansions of $(\beta - z)^{-k}$, are $C_n = \sum_{k=0}^{\infty} \binom{k + n - 1}{n} c_k/\beta^{k+n}$. Thus in the circle $|z| < \beta$

$$f(z) = \sum_{n=0}^{\infty} C_n z^n = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \binom{k + n - 1}{n} z^n(\beta - z)^k c_k/[\beta^n \beta^k(\beta - z)^k],$$

---

[2] I wish to thank Mr. W. Weinstein, who gave me the idea of investigating the continuation of Laurent series.

and writing $1-z/\beta=t$,

$$f(z)=\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\binom{k+n-1}{n}(1-t)^n t^k c_k(\beta-z)^{-k}. \tag{5.2}$$

Thus the continuation of the series (5.1) can be regarded as the generalized sum by series to series transformation by the matrix $A(t)$, where

$$a_{nk}(t)=\binom{k+n-1}{n}(1-t)^n t^k \qquad n,k=0,1,2,\cdots. \tag{5.3}$$

The matrix $A(t)$ is independent of the particular values of z and β, and of the coefficients c_k. The parameter t for all complex values of t defines a family of matrices.

5. I. *$A(t)$ is an α-matrix if and only if $0<t\leq 1$.*

Proof. Obviously $A(0)$ is not an α-matrix, $A(1)$ is an α-matrix. For other values of t the corresponding series to sequence transformation matrix G, given as in (2.4) is $g_{nk}=\sum_{j=0}^{n}\binom{k+j-1}{j}(1-t)^j t^k$, and a short calculation gives

$$g_{nk}-g_{n\,k+1}=\binom{k+n}{n}(1-t)^{n+1}t^k, \tag{5.4}$$

so that $S_n\equiv\sum_k |g_{nk}-g_{n\,k+1}|$ converges if and only if $|t|<1$, and then $S_n=[\,|1-t|/(1-|t|)]^{n+1}$. This is bounded if and only if $|1-t|<1-|t|$. The two conditions together require $0<t<1$, and then $g_{nk}\to 1$ as $n\to\infty$ for every k. We also have:

5. II. *$A(t)$, $t\neq 0$, is an α-matrix if and only if it is non-negative.*

5. III. *$A(t)$, $t\neq 0$, is translative to the right.*

Proof. With the notation of (3.11) for $n>0$ we obtain

$$v'_{n+1}-(1-t)v'_n=tv_{n+1} \text{ and } v'_0=tv_0, \tag{5.5}$$

hence writing $v'_0+v'_1+\cdots+v'_n=\sigma'_n$, $v_0+v_1+\cdots+v_n=\sigma_n$,

$$\sigma'_{n+1}-(1-t)\sigma'_n=t\sigma_{n+1}, \qquad (n=0,1,2,\cdots), \tag{5.6}$$

so that $\lim\sigma'_n=s$ implies $\lim\sigma_n=s$.

5. IV. *$A(t)$ is translative to the left for $|1-t|<1$.*

Proof. From (5.6) follows

$$\sigma'_n = t(1-t)^n\sigma_0 + t(1-t)^{n-1}\sigma_1 + \cdots + t\sigma_n \tag{5.7}$$

and the result follows as in **3. IV.**

Formula (5. 4) shows that the sequence to sequence summation matrix corresponding to $A(t)$ is the matrix $F(t)$, where

$$f_{nk}(t) = \binom{k+n}{n}(1-t)^{n+1}t^k = (1/t-1)a_{n\,k+1}(t). \tag{5.8}$$

5. V. *$F(t)$ is a T-matrix if and only if $0 < t < 1$.*

This follows from (5. 4) and (5. 5), and $F(1)$ is the zero matrix.

5. VI. *Excluding $t = 0$ and $t = 1$, $F(t)$ is translative to the right (left) if and only if $A(t)$ is the same.*

The proof follows from (5. 8) applying (5. 5) to the sequences $0, 0, s_0, s_1, \cdots$ and $0, s_0, s_1, \cdots$ respectively, where σ_n now denotes the $F(t)$ transform of the sequence s_k. The result then follows as in **5. III** and **5. IV.**

5. VII. *$A(t)$ includes $F(t)$ for $t \neq 1$.*

For $t = 0$ the result is trivial. When $t \neq 0$ and the sequence s_k is summable $F(t)$, it follows from **5. III** and **5. VI** that s_{k+1} is summable $F(t)$, so that, by (5. 8),

$$\sum_k f_{nk}(t)s_{k+1} = (1/t-1)a_{n\,k+1}(t)s_{k+1}$$

is convergent for all n.

The result then follows as in **3. X.** We also have as in **3. XI**:

5. VIII. *A necessary condition that $F(t)$ should include $A(t)$ is that $|t| < 1$.*

5. IX. *If the Taylor series $\sum c_k z^k$ has non-zero radius of convergence, and if for $|t| < 1$ $A(t)$ sums the series in a domain D, then $F(t)$ sums the series to the same sum in D*

Corollary. *For any t, $F(t)$ includes $A(t)$ on or outside the circle of convergence in D.* The proof is the same as of **3. XII.**

5. X. *$A(t)$ sums the series $1 + z + z^2 + \cdots$ for every t at $z = 0$, and for $t \neq 0$ if and only if z satisfies simultaneously*

$$|z| < 1/|t|, \tag{5.9}$$

and

(5.10) $$|z - 1/t| > |1 - 1/t|,$$

and the sum is the right value $(1 - z)^{-1}$.

Proof. Applying $A(t)$ to the series we obtain $v_0(z) = (1 - tz)^{-1}$, and for $n > 0$, $v_n(z) = (1 - t)^n tz(1 - tz)^{-n-1}$ provided that $z = 0$ or $|tz| < 1$. Then $\sum v_n(z) = (1 - z)^{-1}$ and only if $z = 0$ or $|(1 - t)/(1 - tz)| < 1$. This concludes the proof.

Thus the method is efficient at $z = 0$ and in a crescent inside the circle (5.9) and outside the circle (5.10). The circle (5.9) has its centre at the origin and passes through $z = 1/t$, the circle (5.10) has its centre at $z = 1/t$ and passes through $z = 1$. *Denoting the domain* (5.9) (5.10) *by* $D(t)$, the origin belongs to $D(t)$ if the real part of $1/t$ is greater than $\frac{1}{2}$, i. e. if $|1 - t| < 1$. The circle (5.9) is inside (5.10), hence $D(t)$ is empty, if $|1 - t| > 2$. In this case $A(t)$ is efficient at the single point $z = 0$ only. If $1 < |1 - t| < 2$, $A(t)$ is efficient in $D(t)$ and at the isolated point $z = 0$. If $-1 < t < 0$, $D(t)$ has no point common with the circle of convergence.

Of particular interest are the methods $A(t)$, for which the origin is in $D(t)$, i. e. for which $|1 - t| < 1$. The domain $D(t)$ is *not a star-domain* as defined in (3.23), since the half-ray from the origin to an intersection of the two circles meets the circle (5.10) again inside the circle (5.9). For a suitably small $|t|$, varying the centre of (5.10) so that $|1 - t| < 1$, $D(t)$ can be made to include any point but $z = 1$. Hence *the union of domains* $D(t)$ *with* $|1 - t| < 1$ *is the whole z-plane with* $z = 1$ *and* $z = \infty$ *excluded.*

The α-matrices $A(t)$, i. e. for which $0 < t \leqq 1$, form a subclass of the above class. The union of domains $D(t)$ of this subclass has as boundary the locus of intersections of the circles (5.9) and (5.10). In polar coordinates with respect to $z = 1$ this boundary is $\rho = (8\cos^3\theta - 4\cos\theta)/(1 - 4\cos^2\theta)$, $\pi/4 \leq \theta \leq \pi/3$, with asymptotes $\theta = \pm\pi/3$, the two branches meeting at $z = 1$, where the tangents make angles $\pm\pi/4$ with the real axis. Thus the union of regular $A(t)$ matrices is efficient in a domain which extends beyond the Borel half-plane.

(5.11) *Definitions.* Let the series $\sum c_k z^k$ have a finite non-zero radius of convergence. Let L be an open Jordan curve joining $z = 0$ and $z = 1$. We denote by ζL the curve formed from the points $\zeta z'$, when z' are all points of L. If the series can be continued along ζL but not at ζ, we say that ζ is a *vertex of the curvilinear star-domain* $S(L)$, and the function $f(z)$, defined by analytic continuations along all curves zL for which z is not a vertex, is said to be defined in the curvilinear star-domain $S(L)$.

The domain $D(t)$ for $|t| < 1$ and $|1-t| < 1$ can be divided into two symmetrical halves each of which is a curvilinear star-domain corresponding to symmetrical circular arcs L and L'. *The partial curvilinear star-domain of $f(z)$ formed with $D(t)$* is the common part of all domains $\zeta D(t)$ for which ζ is a vertex of either of the curvilinear star-domains $S(L)$ or $S(L')$ and for which ζ is in the circle $|\zeta| < Rr''/r'$ (cf. 3.23)).

5. XI. *If $|t| < 1$ and $|1-t| < 1$, $A(t)$ sums the series $\sum c_k z^k$ in the partial curvilinear star-domain formed with $D(t)$, and the sum is the value of $f(z)$ obtained by continuation in this domain.*

Proof. For a fixed t and a regular point z of $f(z)$ we form the series of functions of the complex variable v

$$(5.12) \qquad \sum_{n=0}^{\infty} \phi_n(v) \equiv \sum_{n=0}^{\infty} tz(1-t)^{-1}v^{-2}f(v)[(1-t)/(1-tz/v)]^{n+1}.$$

where $\phi_n(v)$ can be represented by the series

$$(5.13) \qquad \phi_n(v) = tz(1-t)^{-1}v^{-2}f(v) \sum_{k=0}^{\infty} \binom{n+k}{n} (1-t)^{n+1}(tz/v)^k.$$

Both series converge in the domain given by

$$(5.14) \qquad \text{(i)} \quad |v-tz|/|v| > |1-t|, \qquad \text{(ii)} \quad |v| > |tz|,$$

and which we shall denote by $V(t)$.

Here (i) represents the outside of a circle surrounding $v = tz$, the origin being outside and $v = z$ on this circle, while (ii) represents the outside of a circle with centre at the origin and passing through $v = tz$.

(5.15) We suppose that all singularities of $f(v)$ are in $V(t)$.

Constructing three closed Jordan curves C_j $(j = 1, 2, 3)$ from arcs concentric with the boundary of $V(t)$ such that C_i is inside C_{j+1} and all singularities of $f(v)$ outside C_3, we have in the ring between C_1 and C_3

$$(5.16) \qquad \sum_{n=0}^{\infty} \phi_n(v) = zf(v)/v(v-z),$$

and

$$(5.17) \qquad (1/2\pi i)\int_{C_2} \{\sum_{n=0}^{\infty} \phi_n(v)\} dv = f(z) - f(0).$$

Substituting for $\phi_n(v)$ the series (5.13), and integrating term by term the double series, which is uniformly convergent in the ring between C_1 and C_3, we obtain

$$(5.18)\qquad \sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\binom{n+k}{n}(1-t)^n t^{k+1}c_{k+1}z^{k+1}=f(z)-c_0,$$

or rewritten

$$(5.19)\qquad \sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\binom{n+k-1}{n}(1-t)^n t^k c_k z^k=f(z).$$

This was obtained under the supposition (5.15), hence if $v=\zeta$ is a singularity of $f(v)$, (5.14) is satisfied for $v=\zeta$, i.e.

$$(5.20)\qquad \text{(i)}\quad |z/\zeta-1/t|>|1-1/t|,\qquad \text{(ii)}\quad |z/\zeta|<|1/t|.$$

Comparison with (5.9) and (5.10) shows that (5.20) requires z/ζ to be in $D(t)$, i.e. *z is in $\zeta D(t)$*. This proves the theorem.

COROLLARY. *Under the conditions of the theorem $F(t)$ sums the series to the same value.* This follows from **5.IX.**

(5.21) *Example.* The series $0+z/1+z^2/2+\cdots$ represents the function $f(z)\equiv-\log(1-z)$. Applying $A(t)$ with $|1-t|<1$, we obtain for z in $D(t)$

$$(5.22)\qquad \sum v_n(z)=-\log[1-(1-t)/(1-tz)]+\log t-\log(1-tz),$$

where $\log(1+u)$ means the value of the function defined by the convergent series $u+u^2/2+\cdots$. Hence the sum is $-\log(1-z)+2m\pi i$ ($m=0$, 1 or -1) m depending on t and z.

If $t=(1-2i)/5$, $D(t)$ contains points $1<z<5^{\frac{1}{2}}$, so that $z=2$ is in $D(t)$, and for this value of z, $\sum v_n(z)=-\pi i$, which is the value of $f(z)$ when the series is continued in $D(t)$. If t is the conjugate of the previous value, the generalized sum at $z=2$ is $+\pi i$, corresponding to the continuation in $D(t)$. If $z=-2$, which is in $D(t)$ for both values of t, we obtain in both cases $\sum v_n(z)=-\log 3$.

To conclude, we show the connection between the matrices $A(\lambda)$ of (3.5) and $A(t)$ of (5.3). Extending the definition of $\binom{k}{n}$ to negative integral k:

$$\binom{-k}{n}=(-k)(-k-1)\cdots(-k-n+1)/n!=(-1)^n\binom{k+n-1}{n},$$

$$\binom{-k}{0}=1,$$

and substituting $1/t$ for λ, (3.5) gives

$$a_{n\,-k}(\lambda)=\binom{-k}{n}t^{n+k}(1-1/t)^n=\binom{k+n-1}{n}(1-t)^n t^k=a_{nk}(t).$$

I wish to express my thanks to Professor P. Dienes and Dr. R. G. Cooke for their kindness in reading the manuscript and making some useful comments on it. I am also indebted to Birkbeck College for the award of the Armitage-Smith Memorial Prize for 1945/46, during the tenure of which the main part of this work was done.

BIRKBECK COLLEGE,
UNIVERSITY OF LONDON, ENGLAND.

REFERENCES.

R. P. Agnew, "Euler transformations," *American Journal of Mathematics*, vol. 66 (1944), pp. 313-338.

R. G. Cooke, 1. Lectures at *Birkbeck College, University of London.*
2. *Infinite Matrices and Sequence Spaces*, in the press: Macmillan & Co., Ltd.

P. Dienes, *The Taylor Series*, Oxford, 1931.

J. D. Hill, "Some properties of summability," *Duke Mathematical Journal*, vol. 9 (1942), pp. 373-381.

O. Perron, "Zur Theorie der divergenten Reihen," *Mathematische Zeitschrift*, vol. 6 (1920), pp. 158-160 and 286-310.

P. Vermes, 1. "Product of a T-matrix and a γ-matrix," *Journal of the London Mathematical Society*, vol. 21 (1946), pp. 129-134.
2. "On γ-matrices and their application to the binomial series," *Proceedings of the Edinburgh Mathematical Society*, ser. 3, vol. 8 (1947), pp. 1-13.
3. "The application of γ-matrices to Taylor series," the same *Proceedings*, ser. 2, vol. 8 (1948), pp. 43-49.
4. "Gamma matrices and their application to infinite series," *Thesis*, accepted for the award of the Ph. D. degree of the *University of London* (1947).

A CLASS OF INTEGRO-DIFFERENTIAL EQUATIONS.*[1]

By SHIH-HSUN CHANG.

1. **Introduction.** In the present paper, I shall consider equations of the form:

$$\text{(A)} \qquad \partial\phi(x,t)/\partial t = \int_a^b K(x,y) f(y,t,\phi(y,t))\,dy$$

and

$$\text{(B)} \qquad \partial\phi(x,t)/\partial t + p(t)\phi(x,t) = q(x,t) + \int_a^b K(x,y) f(y,t,\phi(y,t))\,dy.$$

It will be assumed that the kernel $K(x,y)$ is such that

$$\int_a^b K^2(x,y)\,dy$$

is a bounded function of x in $a \leq x \leq b$, that $f(x,t,\phi)$ is a continuous function of (x,t,ϕ) for $a \leq x \leq b$, t and ϕ being allowed to take all real values, and that $p(t)$ and $q(x,t)$ are continuous functions of t and of (x,t) respectively. The unknown function to be determined is $\phi(x,t)$.

The equation

$$\text{(A}_0\text{)} \qquad \partial\phi(x,t)/\partial t = \int_a^b K(x,y)\phi(y,t)\,dy,$$

which is a particular case of (A), was considered by Volterra[2] in connection with the solution of linear functional derivative equations. He showed that the unique solution of (A_0) which reduces to a given continuous function $\phi_0(x)$ when $t = t_0$ is given by

$$\phi(x,t) = \phi_0(x) + \int_a^b L(x,y,t)\phi_0(y)\,dy,$$

where

$$L(x,y,t) = \sum_{n=1}^{\infty} ((t-t_0)^n/n!) K_n(x,y),$$

the functions $K_n(x,y)$ being the iterated kernels of $K(x,y)$. It is sometimes convenient to be able to express the solutions of such equations in terms of

---

* Received December 30, 1947.

[1] My warmest thanks are due to Dr. F. Smithies, whose encouragement and advice have been of the utmost value to me in the course of this work.

[2] V. Volterra, (1) p. 394.

the characteristic functions and characteristic values of $K(x,y)$. Barnett[3] has obtained such a result for the special equation (A_0) when $K(x,y)$ is symmetric or skew-symmetric.

In the present paper I shall express the solutions of (A) and (B), for a general kernel $K(x,y)$, in terms of singular functions and singular values of the kernel; the results will include in particular a solution of (A_0) for a general kernel similar to Barnett's solution for symmetric and skew-symmetric kernels.

We shall require some results about the theory of infinite systems of non-linear differential equations which, in order to save space, we assume the reader can easily find in Wintner's papers.[4]

2. Solution of Equation (A). We now discuss the solution of equations of the form (A). We shall suppose that $K(x,y)$ is a real L^2 kernel defined for $a \leq x \leq b$, $a \leq y \leq b$, such that

$$\int_a^b K^2(x,y)\,dy \tag{1}$$

is a bounded function of x, that $\{\lambda_h\}$ is its system of singular values, and $\{\phi_h(x), \psi_h(y)\}$ is its complete orthonormal system of adjoint pairs of singular functions, so that

$$\begin{aligned} \phi_h(x) &= \lambda_h \int_a^b K(x,y)\psi_h(y)\,dy, \\ \psi_h(x) &= \lambda_h \int_a^b K(x,y)\phi_h(y)\,dy. \end{aligned} \tag{2}$$

We write $K(x,y) = K[\phi_h(x), \psi_h(y), \lambda_h]$.

THEOREM 1. *Let $K(x,y)$ satisfy the above conditions and let $f(x,t,u)$ be a continuous function of (x,t,u) for $a \leq x \leq b$ and all real values of t and u such that:*

$$\Phi_i(t, z_1, z_2, \cdots) = (1/\lambda_i)\int_a^b f(y, t, \sum_{h=1}^{\infty} z_h\phi_h(y))\psi_i(y)\,dy \tag{3}$$

is for all values of i a regular power series[5] *in the variables $z_0 = t, z_1, z_2, \cdots$, in the domain σ_r defined by*

$$\sum_{k=0}^{\infty} |z_k|^2 < r^2. \tag{4}$$

---

[3] I. A. Barnett, (2).

[4] A. Wintner, (3), (4), (5), (6).

[5] A. Wintner, (6), p. 242.

Suppose also that for all

$$\boldsymbol{z} = (z_0, z_1, z_2, \cdots)$$

in σ_r *we have*

$$\int_a^b [f(y, t, \sum_{h=1}^{m} z_h \phi_h(y))]^2 dy < A^2 \tag{5}$$

for some constant A *independent of* m *and* t. *Then the integro-differential equation:*

$$\partial\phi(x, t)/\partial t = \int_a^b K(x, y) f(y, t, \phi(y, t)) dy \tag{A}$$

possesses a solution $\phi(x, t)$ *satisfying the initial condition:*

$$\phi(x, 0) = 0 \tag{6}$$

and given by the formula

$$\phi(x, t) = \sum_{i=1}^{\infty} \phi_i(x) z_i(t), \tag{7}$$

where $z_i(t)$ *satisfies the system of equations:*

$$\begin{cases} dz_0/dt = \Phi_0 = 1, \\ dz_i/dt = \Phi_i(\boldsymbol{z}) = \Phi_i(z_0, z_1, z_2, \cdots) \qquad (i = 1, 2, \cdots), \\ z_i(0) = 0. \end{cases} \tag{8}$$

If, in addition, we have

$$| f(x, t, \bar{u}) - f(x, t, u) | \leq C \, | \bar{u} - u | \tag{9}$$

for some constant C, *the solution of* (A) *so obtained is unique.*

By Bessel's inequality applied to (3), we have for all $\boldsymbol{z}$ in σ_r

$$\sum_{i=1}^{\infty} | \lambda_i \Phi_i |^2 \leq \int_a^b [f(y, t, \sum_{h=1}^{\infty} z_h \phi_h(y))]^2 dy < A^2.$$

We suppose the singular values $\{\lambda_h\}$, which are all real, arranged so that $0 < | \lambda_1 | \leq | \lambda_2 | \leq \cdots$; then $| \lambda_h | \to \infty$ as $h \to \infty$. Hence the sequence $\{\Phi_i\}$ is bounded[5] in σ_r i. e.,

$$\sum_{i=1}^{\infty} | \Phi_i(\boldsymbol{z}) |^2 < B < +\infty, \text{ for } \sum_{i=0}^{\infty} | z_i |^2 < r^2. \tag{10}$$

By Wintner's existence theorem,[6] the system of equations (8) possesses a solution such that $\sum_{i=0}^{\infty} | z_i |^2 < +\infty$ provided that

$$| t | < r/2[\boldsymbol{\Phi}]_r,$$

---

[6] A. Wintner, (6), p. 251.

where $[\mathbf{\Phi}]_r$ is the least value of B satisfying (10). We therefore have

$$(11) \qquad \begin{cases} dz_i/dt = (1/\lambda_i) \displaystyle\int_a^b f(y, t, \sum_{h=1}^{\infty} z_h \phi_h(y)) \psi_i(y)\, dy, \\ z_i(0) = 0, \end{cases}$$

for $i = 1, 2, \cdots$.

We now distinguish two cases. Suppose first that $K(x, y)$ has only a finite number of singular values, so that we may write:

$$K(x, y) = \sum_{h=1}^{m} \phi_h(x) \psi_h(y)/\lambda_h.$$

We construct the functions

$$\phi_m(x, t) = \sum_{i=1}^{m} \phi_i(x) z_i(t).$$

We then have, by (11),

$$\partial\phi_m(x, t)/\partial t = \sum_{i=1}^{m} \phi_i(x)\,(dz_i/dt) = \sum_{i=1}^{m} (\phi_i(x)/\lambda_i) \int_a^b f(y, t, \phi_m(y, t)) \psi_i(y) dy$$
$$= \int_a^b K(x, y) f(y, t, \phi_m(y, t))\, dy.$$

Thus $\phi_m(x, t)$ satisfies the integro-differential equation (A) together with the initial condition (6).

Now suppose that $K(x, y)$ has an infinity of singular values. We shall prove that

$$(12) \qquad \lim_{m\to\infty} \phi_m(x, t) = \sum_{i=1}^{\infty} \phi_i(x) z_i(t) = \phi(x, t)$$

exists, and that

$$(13) \qquad \partial\phi(x, t)/\partial t = \sum_{i=1}^{\infty} \phi_i(x)\,(dz_i/dt) = \lim_{m\to\infty} \partial\phi_m(x, t)/\partial t.$$

The results will then follow.

Applying Bessel's inequality to (11), we obtain

$$\sum_{i=1}^{\infty} \lambda_i^2 [z_i'(t)]^2 \leq \int_a^b [f(y, t, \sum_{h=1}^{\infty} z_h \phi_h(y))]^2 dy < A^2.$$

Hence, by Cauchy's inequality, we have:

$$\sum_{i=m+1}^{\infty} | \phi_i(x) z_i'(t) | \leq (\sum_{i=m+1}^{\infty} \phi_i^2(x)/\lambda_i^2)^{\frac{1}{2}} (\sum_{i=m+1}^{\infty} \lambda_i^2 [z_i'(t)]^2)^{\frac{1}{2}}$$
$$< A [K_*(x, x)]^{\frac{1}{2}} \leq AM,$$

where

$$K_*(x, y) = \int_a^b K(x, s) K(y, s)\, ds = \sum_{i=1}^{\infty} \phi_i(x) \phi_i(y)/\lambda_i^2,$$

and M^2 is the upper bound of $K_*(x, x)$ in $a \leq x \leq b$. The series $\sum_{i=1}^{\infty} \phi_i(x) z_i'(t)$ is therefore boundedly convergent, and may therefore be integrated term by term with respect to t. The existence of $\phi(x, t)$ and the equations (12) and (13) are therefore proved.

By a well known expansion theorem,[7] we have

$$\begin{aligned}\chi_m(x, t) &= \partial\phi_m(x, t)/\partial t - \int_a^b K(x, y) f(y, t, \phi_m(y, t))\, dy \\ &= \partial\phi_m(x, t)/\partial t - \sum_{i=1}^{\infty} (\phi_i(x)/\lambda_i) \int_a^b f(y, t, \phi_m(y, t)) \psi_i(y)\, dy \\ &= -\sum_{i=m+1}^{\infty} (\phi_i(x)/\lambda_i) \int_a^b f(y, t, \phi_m(y, t)) \psi_i(y)\, dy.\end{aligned}$$

Hence, by Cauchy's and Bessel's inequalities,

$$\begin{aligned}|\chi_m(x, t)|^2 &\leq \sum_{i=m+1}^{\infty} (\phi_i^2(x)/\lambda_i^2) \int_a^b [f(y, t, \phi_m(y, t))]^2 dy \\ &\leq A^2 \sum_{i=m+1}^{\infty} \phi_i^2(x)/\lambda_i^2 \to 0 \text{ as } m \to \infty,\end{aligned}$$

But $\partial\phi_m(x, t)/\partial t \to \partial\phi(x, t)/\partial t$ as $m \to \infty$, and, by dominated convergence and the continuity of $f(y, t, u)$,

$$\int_a^b K(x, y) f(y, t, \phi_m(y, t))\, dy \to \int_a^b K(x, y) f(y, t, \phi(y, t))\, dy$$

as $m \to \infty$. Hence

$$\partial\phi(x, t)/\partial t = \int_a^b K(x, y) f(y, t, \phi(t, t))\, dy.$$

Thus $\phi(x, t)$ is a solution of equation (A), and it obviously satisfies the initial condition (6). This completes the existence proof.

We now come to the uniqueness part of the theorem. Suppose that the equation (A) has two solutions $\phi(x, t)$ and $\bar{\phi}(x, t)$ satisfying the initial condition (6); we then have

$$\partial\phi(x, t)/\partial t - \partial\bar{\phi}(x, t)/\partial t = \int_a^b K(x, y)[f(y, t, \phi(x, t)) - f(y, t, \bar{\phi}(y, t))]\, dy$$

so that

$$\phi(x, t) - \bar{\phi}(x, t) = \int_0^t du \int_a^b K(x, y)[f(y, u, \phi(y, u) - f(y, u, \bar{\phi}(y, u))]\, dy.$$

Hence, by (9),

---

[7] F. Smithies (7), or E. Schmidt (8).

$$(14)\qquad |\phi(x,t)-\bar{\phi}(x,t)| \leq \int_0^t du \int_a^b |K(x,y)|\, C\, |\phi(y,u)-\bar{\phi}(y,u)|\, dy$$
$$\leq CGN\,|t|,$$

where G is the upper bound of $|\phi(y,u)-\bar{\phi}(y,u)|$ for $a \leq y \leq b$, and fixed interval of values of u containing the interval $(0,t)$, and N is the upper bound of

$$\int_a^b |K(x,y)|\, dy$$

for $a \leq x \leq b$. Repeating the argument, we obtain from (14):

$$|\phi(x,t)-\bar{\phi}(x,t)| \leq C^2GN^2\,|t|^2/2!,$$

and we eventually obtain, by induction,

$$\phi(x,t)-\bar{\phi}(x,t)| \leqq (C^nGN^n\,|t|^n/n!) \to 0 \qquad (n\to\infty).$$

Hence $\phi(x,t)$ and $\bar{\phi}(x,t)$ are identically equal.

Corollary 1. *Let* $\boldsymbol{\eta}=(\eta_1,\eta_2,\cdots)$ *be a point of the domain*:

$$\sigma'_{r/2}:\qquad \sum_{k=1}^{\infty}|\eta_k|^2<\tfrac{1}{4}r^2,$$

and let

$$(15)\qquad \phi_0(x)=\sum_{i=1}^{\infty}\eta_i\phi_i(x),$$

the series being convergent in mean square. If the power series

$$\Phi_i(t,\zeta_1+\eta_1,\zeta_2+\eta_2,\cdots)\qquad (i=1,2,3,\cdots)$$

are regular in the domain:

$$\sigma_{r/2}:\qquad |t|^2+\sum_{k=1}^{\infty}|\zeta_k|^2<\tfrac{1}{4}r^2,$$

then the integro-differential equation (A) *has a solution* $\phi(x,t)$ *satisfying the initial condition* $\phi(x,0)=\phi_0(x)$.

We have, by Minkowski's inequality,

$$\sum_{i=1}^{\infty}|\zeta_i+\eta_i|^2<r^2,$$

so that

$$\int_a^b [f(y,t,\sum_{i=1}^{m}(\zeta_i+\eta_i)\phi_i(y))]^2dy<A^2;$$

the vector $\boldsymbol{\Phi}(t,\zeta_1+\eta_1,\zeta_2+\eta_2,\cdots)$ is therefore bounded, and the system of equations

$$d\zeta_i/dt = \Phi_i(t, \zeta_1 + \eta_1, \zeta_2 + \eta_2, \cdots)$$
$$= (1/\lambda_i) \int_a^b f(y, t, \sum_{h=1}^{\infty} (\zeta_h + \eta_h)\phi_h(y))\psi_i(y)\,dy$$

has a solution $\boldsymbol{\zeta}$ satisfying the initial conditions $\boldsymbol{\zeta}(0) = \mathbf{0}$. This gives the required result.

Theorems similar to Theorem 1 can be established for the adjoint integro-differential equation

$$(\mathrm{A}^*) \qquad \partial\psi(x,t)/\partial t = \int_a^b K(y,x) f(y, t, \psi(y,t))\,dy,$$

and for systems of integro-differential equations of the form

$$(\mathrm{A}_n) \qquad \partial\phi_i(x,t)/\partial t = \int_a^b K^{(i)}(x,y) f_i(y, t, \phi_1(y,t), \cdots \phi_n(y,t))\,dy \qquad (i = 1, 2, \cdots)$$

by exactly similar methods.

3. Special cases. The simplest special case is that in which $f(x,t,u) = u$, so that we have to solve the equation

$$(\mathrm{A}_0) \qquad \partial\phi(x,t)/\partial t = \int_a^b K(x,y)\phi(y,t)\,dy$$

with the initial condition

$$(16) \qquad \phi(x,0) = \phi_0(x).$$

Barnett[8] showed that if $K(x,y)$ is symmetric, the solution of the problem is given by

$$(17) \qquad \phi(x,t) = \phi_0(x) + \sum_{i=1}^{\infty} (e^{t/\lambda_i} - 1)\phi_i(x) \int_a^b \phi_i(y)\phi_0(y)\,dy,$$

where $\{\lambda_i\}$ is the set of characteristic values of $K(x,y)$, and $\{\phi_i(x)\}$ is the corresponding complete orthonormal system of characteristic functions. We shall call equation (17) *Barnett's formula.*

Now suppose that $K(x,y)$ is a general kernel, whose set of singular values is $\{\lambda_h\}$ and whose complete orthonormal system of pairs of adjoint singular functions is $\{\phi_h(x), \psi_h(y)\}$. The system of differential equations (8) then reduces to

$$(18) \qquad dz_i/dt = \sum_{j=1}^{\infty} a_{ij} z_j \qquad (i = 1, 2, 3, \cdots),$$

where

---

[8] I. A. Barnett. (2).

$$a_{ij} = (\phi_j, \psi_i)/\lambda_i = (1/\lambda_i)\int_a^b \phi_j(x)\psi_i(x)\,dx. \tag{19}$$

Let $(\eta_1, \eta_2, \cdots)$ be a set of numbers defined as in Corollary 1, and let $\zeta_i(t) = z_i(t) - \eta_i$. Then, by (18),

$$d\zeta_i/dt = \sum_{j=1}^{\infty} a_{ij}(\zeta_j + \eta_j) = \Phi_i(\boldsymbol{\zeta}), \tag{20}$$

say. Since $\sum_{j=1}^{\infty} | a_{ij} |^2 \leq 1/\lambda_i^2$, the series in (20) is uniformly convergent in any domain of the form $\sum_{j=1}^{\infty} | \zeta_j |^2 < r^2$, and satisfies all the conditions for being a regular power series in any such domain. Hence, by Corollary 1, the system of equations (18) has a solution satisfying the initial conditions:

$$z_i(0) = \eta_i \qquad (i = 1, 2, \cdots). \tag{21}$$

We can determine this solution by Wintner's method[9] as follows. We have

$$S_{i1}(t) = t\Phi_i(0, 0, \cdots) = t\sum_{j=1}^{\infty} a_{ij}\eta_j \qquad (i = 1, 2, \cdots).$$

Hence,[10]

$$\begin{aligned} S'_{i2}(t) &= [\Phi_i(S_{11}, S_{21}, \cdots)]_1 = \Phi_i(S_{11}, S_{12}, \cdots) \\ &= \sum_{j=1}^{\infty} a_{ij}(\eta_j + t\sum_{h=1}^{\infty} a_{jh}\eta_h) \\ &= \sum_{j=1}^{\infty} a_{ij}\eta_j + \mathrm{t}\sum_{h=1}^{\infty} a_{ij}{}^{(2)}\eta_j, \end{aligned}$$

where

$$a_{ij}{}^{(2)} = \sum_{h=1}^{\infty} a_{ih}a_{hj};$$

consequently

$$S_{i2}(t) = t\sum_{j=1}^{\infty} a_{ij}\eta_j + (t^2/2!)\sum_{j=1}^{\infty} a_{ij}{}^{(2)}\eta_j.$$

We obtain by induction

$$S_{in}(t) = \sum_{k=1}^{n} (t^k/k!)\sum_{j=1}^{\infty} a_{ij}{}^{(k)}\eta_j,$$

where $a_{ij}{}^{(k)}$ is defined recursively by the equation

---

[9] A. Wintner, (6), p. 251.

[10] Following Wintner's notation, if $g(t) = \sum_{\nu=0}^{\infty} c_\nu t^\nu$, we use the expression $[g(t)]_n$ to denote $\sum_{\nu=0}^{n} c_\nu t^\nu$.

$$a_{ij}^{(k+1)} = \sum_{h=1}^{\infty} a_{ih}^{(k)} a_{hj};$$

and we have

$$a_{ij}^{(\alpha+\beta)} = \sum_{h=1}^{\infty} a_{ih}^{(\alpha)} a_{hj}^{(\beta)}.$$

Thus the solution of (18) satisfying the initial condition (21) is

$$z_i(t) = \sum_{k=0}^{\infty} (t^k/k!) \sum_{j=1}^{\infty} a_{ij}^{(k)} \eta_j \qquad (i = 1, 2, \cdots), \tag{22}$$

where we have taken

$$\left.\begin{aligned} a_{ij}^{(0)} = \delta_{ij} &= 1 \qquad (i = j) \\ &= 0 \qquad (i \neq j) \end{aligned}\right\}.$$

The solution of (A) satisfying the initial condition (16) is therefore

$$\begin{aligned} \phi(x, t) &= \sum_{i=1}^{\infty} \sum_{\rotatebox{180}{$\scriptstyle k=0$}}^{\infty} \sum_{j=1}^{\infty} a_{ij}^{(k)} \phi_i(x) (t^k/k!) \eta_j \\ &= \sum_{i=1}^{\infty} \sum_{k=0}^{\infty} \sum_{j=1}^{\infty} a_{ij}^{(k)} \phi_i(x) (t^k/k!) \int_a^b \phi_0(\xi) \phi_j(\xi) \, d\xi. \end{aligned} \tag{23}$$

Since $f(x, t, u) = u$ satisfies the condition (9) this solution is unique.

Equation (23) contains Barnett's formula as a particular case; for, if $K(x, y)$ is symmetric, we have

$$\begin{aligned} a_{ij}^{(k)} &= 0 \qquad (i \neq j), \\ &= 1/\lambda_i^k \qquad (i = j). \end{aligned}$$

We can use these results to solve the integro-differential equation:

$$\partial\phi(x, t)/\partial t = f(t) \int_a^b K(x, y) \phi(y, t) \, dy, \tag{24}$$

where $f(t)$ is any positive integrable function of t, with the initial condition $\phi(x, 0) = \phi_0(x)$. For, if we change to the new variable:

$$w = \int_0^t f(u) \, du,$$

the equation reduces to one of the form (A_0). We find that (24) has the unique solution

$$\phi(x, t) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \{ \sum_{k=0}^{\infty} (a_{ij}^{(k)}/k!) [\int_0^t f(u) \, du]^k \phi_i(x) \int_a^b \phi_0(\xi) \phi_j(\xi) \, d\xi \}.$$

4. More general equations. In order to solve the more general equation

$$\text{(B)}\quad (\partial\phi(x,t)/\partial t) + p(t)\phi(x,t) = q(x,t) + \int_a^b K(x,y)f(y,t,\phi(y,t))dy,$$

we put

$$w(x,t) = \phi(x,t)\exp\left(\int_0^t p(u)du\right) - \int_0^t q(x,u)\exp\left(\int_0^u p(v)dv\right)du,$$

so that

$$\partial w(x,t)/\partial t = [\partial\phi(x,t)/\partial t + p(t)\cdot\phi(x,t) - q(x,t)]\exp\left(\int_0^t p(u)du\right).$$

Thus, if we multiply both sides of (B) by $\exp\left(\int_0^t p(u)du\right)$, we obtain the equation:

$$(25)\quad \partial w(x,t)/\partial t = \exp\left(\int_0^t p(u)du\right)\cdot\int_a^b K(x,y)f[y,t,\psi(y,t)]dy,$$

where

$$\psi(y,t) = \exp\left(-\int_0^t p(u)du\right)\left[w(y,t) + \int_0^t q(y,u)\exp\left(\int_0^u p(v)dv\right)du\right].$$

This is of the form (A), and can be treated by the methods of **2**.

As a particular case of this, we can show that the equation

$$\partial\phi(x,t)/\partial t = p(t)\phi(x,t) + \int_a^b K(x,y)\phi(y,t)dy,$$

with the initial condition $\phi(x,0) = \phi_0(x)$, has the solution

$$\phi(x,t) = \exp\left(\int_0^t p(u)du\right)\sum_{i=1}^{\infty}\phi_i(x)\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}a_{ij}^{(k)}(t^k/k!)(\phi_0,\phi_j).$$

This is a generalization of another formula of Barnett.[11]

Note. If in **2** we assume that $K(x,y)$ is an L^2 kernel without further restriction, then

$$\int_a^b K^2(x,y)dy$$

is finite for almost all x, and the series $\sum_{i=1}^{\infty}(\phi_i^2(x)/\lambda_i^2)$ converges for almost all x; therefore the series $\sum_{i=1}^{\infty}\phi_i(x)z'_i(t)$ is dominatedly convergent for almost all x, and can be integrated term by term with respect to t for almost all x,

---

[11] I. A. Barnett, (2), p. 201, equation (13).

i. e. the function (12) exists and (13) is true for almost all x. In this case we have:[12]

$$\partial\phi_m(x,t)/\partial t - \int_a^b K(x,y)f(y,t,\phi_m(y,t)\,dy$$

$$= \lim_{n\to\infty} \sum_{i=m+1}^{n} (\phi_i(x)/\lambda_j) \int_a^b f(y,t,\phi_m(y,t)\psi_i(y)\,dy \to 0$$

$$(\text{as } m \to \infty),$$

i. e.

$$\partial\phi(x,t)/\partial t = \int_a^b K(x,y)f(y,t,\phi(y,t))\,dy$$

for almost all x. It is also easy to show that if the equation (A) has two solutions $\phi(x,t)$ and $\phi(x,t)$ satisfying the initial condition (6), then

$$\phi(x,t) = \bar{\phi}(x,t)$$

for almost all x.

FITZWILLIAM HOUSE,
CAMBRIDGE, ENGLAND
AND
SZECHUAN, CHINA.

REFERENCES.

(1) V. Volterra, "Sulle equazioni alle derivate funzionali," *Atti, Reale Accademia dei Lincei*, serie 5, vol. 23 (1914), pp. 393-399.

(2) I. A. Barnett, "Integro-differential equations with constant limits of integration," *Bulletin of the American Mathematical Society*, vol. 26 (1919), pp. 192-203.

(3) A. Wintner, "Zur Theorie der unendlichen Differentialsysteme," *Mathematische Annalen*, vol. 95 (1926), pp. 544-556.

(4) A. Wintner, "Zur Lösung von Differentialsystem mit unendlich vielen Veränderlichen," *Mathematische Annalen*, vol. 98 (1927), pp. 273-280.

(5) A. Wintner, "Zur Analysis im Hilbertschen Raume," *Mathematische Zeitschrift*, vol. 28 (1928), pp. 451-470.

(6) A. Wintner, "Upon a theory of infinite systems of non-linear implicit and differential equations," *American Journal of Mathematics*, vol. 53 (1931), pp. 241-257.

(7) F. Smithies, "The eigen-values and singular values of integral equations," *Proceedings of the London Mathematical Society*, series 2, vol. 43 (1937), pp. 255-279.

(8) E. Schmidt, "Zur Theorie der linearen und nichtlinearen Integralgleichungen," *Mathematische Annalen*, vol. 63 (1907), pp. 433-476.

---

[12] F. Smithies, (7), p. 269.

STRUCTURE OF THE HOMOTOPY GROUPS OF MAPPING SPACES.*

By SZE-TSEN HU.

---

1. Introduction. In 1939, M. Abe [1] established a theorem concerning the structure of the fundamental group of the space of spherical mappings into a metric space. A recent paper of the author [5] solves the structure of the higher homotopy groups of the same space. After the publication of the paper [5], A. D. Wallace proposed to the author the problem to determine the structure of the homotopy groups of a general mapping space. The present work is a consequence of this fruitful suggestion.

Throughout the present paper, let Y denote a pathwise connected topological space and y_0 a given point of Y. Let X be a connected finite geometric simplicial complex; X_0 a closed subcomplex of X (not necessarily connected), and $X_* = X - X_0$ its open complement. Let m denote the dimension of X_*.

Let us consider the totality of the mappings (i. e. continuous transformations) $f : X \rightarrow Y$ with $f(X_0) = y_0$. These mappings form a space Ω with the compact-open topology of R. H. Fox [4], which can be described as follows. For any two sets A in X and W in Y, let $M(A, W)$ denote the set of mappings $f \varepsilon \Omega$ for which $f(A) \subset W$. The compact-open topology of Ω is defined by selecting as a subbase [6, p. 6] for the open sets of Ω the sets $M(A, W)$ where A ranges over the compact subsets of X and W ranges over the open subsets of Y. Let $O : X \rightarrow Y$ denote the constant mapping $O(X) = y_0$. We shall denote by Ω_0 the path-component of Ω containing O, i. e. the set of mappings $f \varepsilon \Omega$ which can be connected to O by a path in Ω. For simplicity, we shall always denote by $\pi_n = \pi_n(Y)$, $n \geqq 1$, the n-th homotopy group of Y with $y_0 \varepsilon Y$ as the base point. Our primary object is to study the structure of the homotopy groups $\pi_r(\Omega_0)$, $(r = 1, 2, \cdots)$, with $O \varepsilon \Omega_0$ as the base point.

Our main results can be briefly summarized as follows. For each $r \geqq 1$, there is a homomorphism (called *the tubular homomorphism*) $\tau_r : \pi_r(\Omega_0) \rightarrow \pi_r(Y)$ such that τ_r is onto if X_0 is empty and τ_r is trivial (i. e. its image consists of a single element) otherwise. The kernel K of τ_r

---

* Received November 8, 1948.

(called *the principal subgroup of* $\pi_r(\Omega_0)$) is a solvable group [7, p. 15] and has a decreasing sequence of subgroups

$$K = K_0 \supset K_1 \supset \cdots \supset K_n \supset \cdots \supset K_m = 1$$

such that each is a normal subgroup of the preceding and that the quotient group K_{n-1}/K_n $(n = 1, 2, \cdots, m)$ is abelian and isomorphic with the difference group

$$H_*^n(X_*, \pi_{n+r}) = P^n(X_*, \pi_{n+r}) - R^n(X_*, \pi_{n+r}),$$

where the two groups displayed on the right member are subgroups of the n-th cohomology group $H^n(X_*, \pi_{n+r})$ of the open subcomplex $X_* = X - X_0$ (which is an abstract complex in the sense of A. W. Tucker [6, p. 89]) with the homotopy group $\pi_{n+r} = \pi_{n+r}(Y)$ as the coefficient group. If X_0 is non-vacuous, then $\pi_r(\Omega_0) = K$; otherwise, there exists a subgroup T (called *the tubular subgroup*) of $\pi_r(\Omega_0)$ which is mapped isomorphically onto $\pi_r(Y)$ by τ_r. Then it follows that: (1) if $r > 1$, $\pi_r(\Omega_0)$ is the direct sum of the subgroups T and K; and (2) $\pi_1(\Omega_0)$ is the direct product of T and K if and only if T is a normal subgroup of $\pi_1(\Omega_0)$.

A part of the above results overlap with an unpublished work of S. Wylie, who has obtained a few theorems concerning the structure of the fundamental group of mapping spaces, by a quite different approach with some geometric methods.

2. Preliminaries.

Let E^r denote the r-dimensional parallelotope in the euclidean r-space defined by

$$E^r\colon\ 0 \leqq e_i \leqq 1 \qquad (i = 1, 2, \cdots, r),$$

where $(e_1, e_2, \cdots, e_r)$ denotes the coordinates of an arbitrary point in the euclidean r-space. Denote by E_1^r and E_2^r the closed subsets of E^r defined respectively by the inequalities $e_1 \leqq \frac{1}{2}$ and $e_1 \geqq \frac{1}{2}$.

An element $\alpha \in \pi_r(Y)$ is represented by a mapping $f\colon E^r \to Y$ which maps the boundary ∂E^r into the point y_0. Two such mappings represent the same element of $\pi_r(Y)$ if and only if they are homotopic relative to ∂E^r. Suppose $\alpha, \beta \in \pi_r(Y)$ be respectively represented by the mappings

$$f, g\colon E^r \to Y, \qquad f(\partial E^r) = y_0 = g(\partial E^r).$$

Then the element $\alpha\beta^{-1} \in \pi_r(Y)$ is represented by the mapping $h\colon E^r \to Y$ defined by

$$h(e_1, e_2, \cdots, e_r) = \begin{cases} f(2e_1, e_2, \cdots, e_r) & (0 \leqq e_1 \leqq \frac{1}{2}) \\ g(2 - 2e_1, e_2, \cdots, e_r) & (\frac{1}{2} \leqq e_1 \leqq 1). \end{cases}$$

Now, let I denote the closed interval $(0, 1)$ of real numbers and let σ denote an arbitrary n-simplex with a fixed orientation. Then the topological product $\Sigma = \sigma \times E^r \times I$ is an oriented $(n + r + 1)$-dimensional geometric cell. Let

$$\Sigma_0 = (\sigma \times \partial E^r \times I) \cup (\sigma \times E^r \times 1).$$

Clearly Σ_0 is contained in the boundary $\partial\Sigma$ of Σ.

Let us consider an arbitrarily given mapping

$$\phi\colon \partial\Sigma \to Y, \qquad \phi(\Sigma_0) = y_0.$$

Since Σ_0 is nonvacuous and connected, ϕ determines a unique element $(\phi, \Sigma) \varepsilon \pi_{n+r}(Y)$ which depends only on the homotopy class $[\phi]$ relative to Σ_0 and the orientation of Σ. For two arbitrary mappings

$$\phi, \psi\colon \partial\Sigma \to Y, \qquad \phi(\Sigma_0) = y_0 = \psi(\Sigma_0),$$

let us define a mapping $\chi\colon \partial\Sigma \to Y$ as follows. For each point

$$(x, e_1, e_2, \cdots, e_r, t) \varepsilon \partial\Sigma \subset \Sigma,$$

where

$$x \varepsilon \sigma, \qquad (e_1, e_2, \cdots, e_r) \varepsilon E^r, \qquad t \varepsilon I,$$

we take

$$\chi(x, e_1, e_2, \cdots, e_r, t) = \begin{cases} \phi(x, 2e_1, e_2, \cdots, e_r, t), & (0 \leqq e_1 \leqq \frac{1}{2}), \\ \psi(x, 2 - 2e_1, e_2, \cdots, e_r, t), & (\frac{1}{2} \leqq e_1 \leqq 1). \end{cases}$$

Then it is clear that $\chi(\Sigma_0) = y_0$ and $(\chi, \Sigma) = (\phi, \Sigma) - (\psi, \Sigma)$. Here we have used the additive notation for the group operation in $\pi_{n+r}(Y)$, because in the sequel we shall consider only the simplexes $\sigma \varepsilon X$ with dimension $n > 0$ and hence $\pi_{n+r}(Y)$ is abelian.

E^r can be considered as a cell complex with an r-cell E^r and a number of cells with dimensions less than r on its boundary. Then the topological product $A = X \times E^r$ is also a cell complex. Let $A_0 = (X \times \partial E^r) \cup (X_0 \times E^r)$, then A_0 is a closed subcomplex of A and its complement $A_* = A - A_0$ an open subcomplex of A.

According to R. H. Fox [4], the mappings $f\colon E^r \to \Omega_0$, $f(\partial E^r) = 0$ are in a (1-1)-correspondence (which preserves homotopy) with the mappings $F\colon A \to Y$, $F(A_0) = y_0$. Hence, an element $\alpha \varepsilon \pi_r(\Omega_0)$ is represented by such a mapping F; and two such mappings represent the same element of

$\pi_r(\Omega_0)$ if and only if they are homotopic relative to A_0. If two arbitrary elements $\alpha, \beta \varepsilon \pi_r(\Omega_0)$ be represented by the mappings $F, G\colon A \to Y$, $F(A_0) = y_0 = G(A_0)$, then the element $\alpha\beta^{-1}$ is represented by the mapping $H\colon A \to Y$ defined for each $x \varepsilon X$ and each $(e_1, e_2, \cdots, e_r) \varepsilon E^r$:

$$H(x, e_1, e_2, \cdots, e_r) = \begin{cases} F(x, 2e_1, e_2, \cdots, e_r), & (o \leqq e_1 \leqq \tfrac{1}{2}), \\ G(x, 2-2e_1, e_2, \cdots, e_r), & (\tfrac{1}{2} \leqq e_1 \leqq 1). \end{cases}$$

In the remainder of the present section, we shall give a trivial generalization of a theorem due to N. E. Steenrod [8, p. 316] which will be used in the sequel. Let G denote an arbitrary abelian group.

(2.1) *The correspondence $\sigma \times E^r \to \sigma$ defines an isomorphic cochain mapping $\iota_r\colon A_* \to X_*$ and thereby induces isomorphisms onto:*

$$\iota_r\colon H^{n+r}(A_*, G) \to H^n(X_*, G), \qquad n \geqq 0.$$

Proof. That ι_r is one-to-one is trivial. Since the coboundary $\delta E^r = 0$, we have $\delta(\sigma \times E^r) = \delta\sigma \times E^r$. Hence ι_r commutes with the coboundary operator δ, i. e. ι_r is a cochain mapping. Q. E. D.

Now, let

$$B = A \times I = X \times E^r \times I,$$

$$B_0 = (A \times o) \cup (A_0 \times I) \cup (A \times 1), \qquad B_* = B - B_0.$$

According to N. E. Steenrod [8, p. 316], the cochain mapping

$$\sigma \times E^r \times I \to \sigma \times E^r$$

induces isomorphisms onto:

$$\iota'\colon H^{n+r+1}(B_*, G) \to H^{n+r}(A_*, G), \qquad n \geqq 0.$$

In the sequel, we shall make use of the combined isomorphisms:

$$\iota = \iota_r\iota'\colon H^{n+r+1}(B_*, G) \to H^n(X_*, G), \qquad n \geqq 0.$$

The notations of this section will be used for the whole paper.

3. Some cohomology invariants. Following N. E. Steenrod [8], we shall put

$$\bar{A}^p = A_0 \cup A^p, \quad \bar{B}^p = B_0 \cup B^p, \qquad (p \geqq 0),$$

where A^p (B^p) denotes the p-dimensional skeleton of A (B), i. e. the set of cells of A (B) with dimensions not exceeding p. Let

$$B_1 = (A_0 \times I) \cup (A \times 1).$$

Consider an arbitrary mapping

$$f\colon \bar{B}^{n+r} \to Y, \qquad f(B_1) = y_0, \qquad\qquad n > 0.$$

For an arbitrary $(n+r+1)$-cell $s_i = \sigma \times E^r \times I \varepsilon B_*$, $\sigma \varepsilon X_*$, clearly we have $f[(\sigma \times \partial E^r \times I) \cup (\sigma \times E^r \times 1)] = y_0$. Hence, according to **2**, $f \mid \partial s_i$ determines a unique element $(f, s_i) \varepsilon \pi_{n+r}(Y)$. By the argument of S. Eilenberg [3, p. 237], the cochain $c^{n+r+1}(f) = \Sigma_i (f, s_i) s_i$ is a cocycle in B_*. The cocycle $c^{n+r+1}(f)$ represents an element $\gamma^{n+r+1}(f)$ of the cohomology group $H^{n+r+1}(B_*, \pi_{n+r})$, which depends only on the homotopy class $[f]$ of f relative to B_1 The element $\gamma^{n+r+1}(f)$ is said to be *presented* by f, and f a *presenter* of $\gamma^{n+r+1}(f)$. Such elements are called *presentable elements* of $H^{n+r+1}(B_*, \pi_{n+r})$. If $\alpha, \beta \varepsilon H^{n+r+1}(B_*, \pi_{n+r})$ be arbitrary presentable elements of $H^{n+r+1}(B_*, \pi_{n+r})$ which are presented by the mappings

$$f, g\colon \bar{B}^{n+r} \to Y, \qquad f(B_1) = y_0 = g(B_1),$$

then the element α-β is clearly presented by the mapping $h\colon \bar{B}^{n+r} \to Y$ defined by

$$h(x, e_1, e_2, \cdots, e_r, t) = \begin{cases} f(x, 2e_1, e_2, \cdots, e_r, t), & (o \leqq e_1 \leqq \tfrac{1}{2}), \\ g(x, 2 - 2e_1, e_2, \cdots, e_r, t), & (\tfrac{1}{2} \leqq e_1 \leqq 1), \end{cases}$$

for each point $(x, e_1, e_2, \cdots, e_r, t) \varepsilon \bar{B}^{n+r} \subset B$, where $x \varepsilon X$, $(e_1, e_2, \cdots, e_r) \varepsilon E^r$ and $t \varepsilon I$. Hence we have proved the following theorem.

(3.1) *The presentable elements of the cohomology group* $H^{n+r+1}(B_*, \pi_{n+r})$ *form a subgroup* $P^{n+r+1}(B_*, \pi_{n+r})$, *called the presentable subgroup of* $H^{n+r+1}(B_*, \pi_{n+r})$.

A mapping $f\colon \bar{B}^{n+r} \to Y$ is said to be *regular* in case $f(B_0) = y_0$. An element of $H^{n+r+1}(B_*, \pi_{n+r})$ is said to be *regular* if it can be presented by a regular mapping. For two arbitrary regular mappings $f, g\colon \bar{B}^{n+r} \to Y$, the mapping h defined above is obviously regular. Hence, we have the following theorem.

(3.2) *The regular elements of the cohomology group* $H^{n+r+1}(B_*, \pi_{n+r})$ *form a subgroup* $R^{n+r+1}(B_*, \pi_{n+r})$, *called the regular subgroup of* $H^{n+r+1}(B_*, \pi_{n+r})$.

Let us denote by $P^n(X_*, \pi_{n+r})$ and $R^n(X_*, \pi_{n+r})$ the images of $P^{n+r+1}(B_*, \pi_{n+r})$ and $R^{n+r+1}(B_*, \pi_{n+r})$ under the isomorphism

$$\iota\colon H^{n+r+1}(B_*, \pi_{n+r}) \to H^n(X_*, \pi_{n+r})$$

described in 2, and define

$$H_\bullet^n(X_\bullet, \pi_{n+r}) = P^n(X_\bullet, \pi_{n+r}) - R^n(X_\bullet, \pi_{n+r})$$

for each $n = 1, 2, \cdots, m$. The groups $P^n(X_\bullet, \pi_{n+r})$ and $R^n(X_\bullet, \pi_{n+r})$ are respectively called the *presentable* and the *regular subgroup* of the cohomology group $H^n(X_\bullet, \pi_{n+r})$, and their elements are called the *presentable* and the *regular elements* of the same.

(3.3) *If* $H^{p-1}(X_\bullet, \pi_{p+r}) = 0$ *for each* $1 < p < n$, *then* $H_\bullet^n(X_\bullet, \pi_{n+r})$ *is isomorphic with* $P^n(X_\bullet, \pi_{n+r})$.

Proof. We need only to prove that $R^{n+r+1}(B_\bullet, \pi_{n+r}) = 0$. Let α be an arbitrary regular element of $H^{n+r+1}(B_\bullet, \pi_{n+r})$; then, by definition, there exists a mapping $f\colon \bar{B}^{n+r} \to Y$ such that $f(B_0) = y_0$ and $\gamma^{n+r+1}(f) = \alpha$. Every $(r+1)$-cell of $B_\bullet$ is of the form $s_i = x_i \times E^r \times I$, where x_i is a vertex of $X_\bullet$. Since the boundary ∂s_i of s_i is in B_0, the partial mapping $f \mid s_i$ defines an element $\beta_i \varepsilon \pi_{r+1}(Y)$. Since X is connected, all the β_i are equal to a fixed $\beta \varepsilon \pi_{r+1}(Y)$. Clearly $\beta = 0$ if X_0 is nonvacuous. Choose a representative

$$\phi\colon E^r \times I \to Y, \qquad \phi[(E^r \times o) \cup (\partial E^r \times I) \cup (E^r \times 1)] = y_0,$$

for the element β.

Define a mapping $g^*\colon B \to Y$ by taking

$$g^*(x, e, t) = \phi(e, t), \qquad (x \varepsilon X, \qquad e \varepsilon E^r, \qquad t \varepsilon I).$$

Let $g = g^* \mid \bar{B}^{n+r}$. Since g has g^* as an extension over B, we have $\gamma^{n+r+1}(g) = 0$. Define a mapping $h\colon \bar{B}^{n+r} \to Y$ by taking

$$h(x, e_1, e_2, \cdots, e_r, t) = \begin{cases} f(x, 2e_1, e_2, \cdots, e_r, t), & (o \leqq e_1 \leqq \tfrac{1}{2}), \\ g(x, 2 - 2e_1, e_2, \cdots, e_r, t), & (\tfrac{1}{2} \leqq e_1 \leqq 1), \end{cases}$$

for each point $(x, e_1, e_2, \cdots, e_r, t) \varepsilon \bar{B}^{n+r} \subset B$, where $x \varepsilon X$, $(e_1, e_2, \cdots, e_r) \varepsilon E^r$, and $t \varepsilon I$. Clearly we have $h(B_0) = y_0$ and

$$\gamma^{n+r+1}(h) = \gamma^{n+r+1}(f) - \gamma^{n+r+1}(g) = \alpha.$$

Now, for each $(r+1)$-cell $s_i \varepsilon B_\bullet$, the partial mapping $h \mid s_i$ represents the element $\beta - \beta = 0$ of the homotopy group $\pi_{r+1}(Y)$. Hence there exists a homotopy $h_t\colon \bar{B}^{n+r} \to Y$ $(o \leqq t \leqq 1)$ such that $h_0 = h$, $h_1(\bar{B}^{r+1}) = y_0$, and $h_t(B_0) = y_0$ for each $o \leqq t \leqq 1$.

Since, for each $p = 2, 3, \cdots, n-1$, we have

$$H^{p+r}(B_\bullet, \pi_{p+r}) \approx H^{p-1}(X_\bullet, \pi_{p+r}) = 0,$$

it follows from the successive application of the first homotopy theorem of S. Eilenberg [3, p. 240] that there exists a homotopy $k_t\colon \bar{B}^{n+r} \to Y$ $(o \leqq t \leqq 1)$ such that $k_0 = h_1$, $k_1(\bar{B}^{n+r-1}) = y_0$, and $k_t(B_0) = y_0$ for each $o \leqq t \leqq 1$. Since $k_1 \mid \bar{B}^{n+r-1}$ can be extended throughout B, it follows from the first extension theorem of S. Eilenberg [3, p. 239] that

$$\alpha = \gamma^{n+r+1}(h) = \gamma^{n+r+1}(k_0) = \gamma^{n+r+1}(k_1) = 0.$$

Hence $R^{n+r+1}(B_*, \pi_{n+r}) = 0$. This completes the proof.

(3.4) *If* $H^{p+1}(X_*, \pi_{p+r}) = 0$ *for each* $n < p < m$, *then* $P^n(X_*, \pi_{n+r})$ *coincides with* $H^n(X_*, \pi_{n+r})$.

Proof. We need only to prove that $P^{n+r+1}(B_*, \pi_{n+r}) = H^{n+r+1}(B_*, \pi_{n+r})$. Let c be an arbitrary cocycle in B_* with coefficients in π_{n+r}. It can be easily seen that there exists a mapping

$$F_0\colon (\bar{A}^{n+r} \times o) \cup (\bar{A}^{n+r-1} \times I) \cup (A \times 1) \to Y$$

such that

$$F_0[(\bar{A}^{n+r-1} \times I) \cup (A \times 1)] = y_0, \qquad c^{n+r+1}(F_0) = c.$$

Define a mapping $f_0\colon \bar{A}^{n+r} \to Y$ by taking $f_0(a) = F_0(a, o)$, $(a \,\varepsilon\, \bar{A}^{n+r})$. Since c is a cocycle in B_*, f_0 admits an extension $f_1\colon \bar{A}^{n+r+1} \to Y$. f_1 determines an element $\gamma^{n+r+2}(f_1) \,\varepsilon\, H^{n+r+2}(A_*, \pi_{n+r+1})$. Since

$$H^{n+r+2}(A_*, \pi_{n+r+1}) \approx H^{n+2}(X_*, \pi_{n+r+1}) = 0,$$

we have $\gamma^{n+r+2}(f_1) = 0$. It follows from the first extension theorem of S. Eilenberg [3, p. 239] that f_0 has an extension $f_2\colon \bar{A}^{n+r+2} \to Y$. By successive application of this argument, one can prove that f_0 admits an extension $f\colon A \to Y$. Extend F_0 to a mapping $F\colon \bar{B}^{n+r} \to Y$ by taking $F(a, o) = f(a)$, $(a \,\varepsilon\, A)$. Hence we have $c = {}^{n+r+1}(F)$, i.e. every element of the group $H^{n+r+1}(B_*, \pi_{n+r})$ is presentable. This completes the proof.

4. The *n*-trivial subgroup. We have seen in **2** that the elements of the homotopy group $\pi_r(\Omega_0)$ are the homotopy classes relative to A_0 of the mappings $f\colon A \to Y$ with $f(A_0) = y_0$.

An element $\alpha \,\varepsilon\, \pi_r(\Omega_0)$ is said to be *n-trivial* $(n \geqq o)$ if it can be represented by a mapping $f\colon A \to Y$ such that $f(\bar{A}^{n+r}) = y_0$. The n-trivial elements of $\pi_r(\Omega_0)$ obviously form a subgroup K_n of $\pi_r(\Omega_0)$, called the *n-trivial subgroup* of $\pi_r(\Omega_0)$. Thus, we have a decreasing sequence of subgroups of $\pi_r(\Omega_0)$:

$$\pi_r(\Omega_\sigma) \supset K_0 \supset \cdots \supset K_{n-1} \supset K_n \supset \cdots \supset K_m,$$

where the last group K_m clearly consists of a single element.

Let $\alpha \varepsilon K_{n-1}$ $(n > o)$ be an arbitrary $(n-1)$-trivial element of $\pi_r(\Omega_0)$ and let $f\colon A \to Y$, $f(\bar{A}^{n+r-1}) = y_0$, be a representative of α. Define a mapping $F\colon \bar{B}^{n+r} \to Y$ by taking

$$F(a, t) = \begin{cases} f(a) & (a \varepsilon A, \quad t = 0) \\ y_0 & (a \varepsilon \bar{A}^{n+r-1}, \quad t \varepsilon I) \\ y_0 & (a \varepsilon A, \quad t = 1). \end{cases}$$

Since $F(B_1) = y_0$, F presents a presentable element

$$\gamma^{n+r+1}(F) \varepsilon P^{n+r+1}(B_*, \pi_{n+r}).$$

Next, let $g\colon A \to Y$, $g(\bar{A}^{n+r-1}) = y_0$, be another representative of α, and $G\colon \bar{B}^{n+r} \to Y$ be similarly defined for g. Define a mapping $\Phi\colon \bar{B}^{n+r} \to Y$ by taking

$$\Phi(x, e_1, e_2, \cdots, e_r, t) = \begin{cases} F(x, 2e_1, e_2, \cdots, e_r, t) & (o \leqq e_1 \leqq \tfrac{1}{2}), \\ G(x, 2 - 2e_1, e_2, \cdots, e_r, t) & (\tfrac{1}{2} \leqq e_1 \leqq 1), \end{cases}$$

for each point $(x, e_1, e_2, \cdots, e_r, t) \varepsilon \bar{B}^{n+r} \subset B$, where $x \varepsilon X$, $(e_1, e_2, \cdots, e_r) \varepsilon E^r$, and $t \varepsilon I$. Clearly $\gamma^{n+r+1}(\Phi) = \gamma^{n+r+1}(F) - \gamma^{n+r+1}(G)$. Define a mapping $\phi\colon A \to Y$ by taking $\phi(a) = \Phi(a, o)$, $(a \varepsilon A)$. Since ϕ represents the element $\alpha\alpha^{-1} = 1$ of $\pi_r(\Omega_0)$, there exists a homotopy $\phi_t\colon A \to Y$, $(o \leqq t \leqq 1)$, such that $\phi_0 = \phi$, $\phi_1(A) = y_0$, and $\phi_t(A_0) = y_0$ for each $o \leqq t \leqq 1$.

Define a homotopy $\Phi_\lambda\colon B_0 \to Y$ $(o \leqq \lambda \leqq 1)$ by taking

$$\Phi_\lambda(a, t) = \begin{cases} \phi_\lambda(a) & (a \varepsilon A, \quad t = 0) \\ y_0 & (a \varepsilon A_0, \quad t \varepsilon I) \\ y_0 & (a \varepsilon A, \quad t = 1). \end{cases}$$

Since $\Phi_0 = \Phi$ is defined over $\bar{B}^{n+r}$, it follows from the homotopy extension property [2, p. 501] that Φ_λ has an extension $\Phi_\lambda^*\colon \bar{B}^{n+r} \to Y$, $(o \leqq \lambda \leqq 1)$, such that $\Phi_0^* = \Phi$. Since $\Phi_1^*(B_0) = y_0$, hence $\gamma^{n+r+1}(\Phi) = \gamma^{n+r+1}(\Phi_0^*)$ is a regular element, $\gamma^{n+r+1}(F)$ determines an element of $H_*^n(X_*, \pi_{n+r})$, which depends only on α and will be denoted by $\mu_n(\alpha)$.

(4.1) *The correspondence* $\alpha \to \mu_n(\alpha)$ *is a homomorphism*

$$\mu_n\colon K_{n-1} \to H_*^n(X_*, \pi_{n+r})$$

of the $(n-1)$*-trivial subgroup* $K_{n-1} \subset \pi_r(\Omega_0)$ *onto the group* $H_*^n(X_*, \pi_{n+r})$, *whose kernel is the n-trivial subgroup* K_n.

Proof. That μ_n is a homomorphism is clear. To prove that μ_n is onto, let $\xi \in P^{n+r+1}(B_*, \pi_{n+r})$ be arbitrarily given. By definition, there exists a mapping $F\colon \bar{B}^{n+r} \to Y$, $F(B_1) = y_0$, such that $\gamma^{n+r+1}(F) = \xi$. Define a homotopy

$$F_\lambda\colon (\bar{A}^{n+r-1} \times I) \cup (A \times 1) \to Y \qquad (o \leqq \lambda \leqq 1)$$

by taking

$$F_\lambda(a, t) = \begin{cases} F(a, \lambda + t - \lambda t) & (a \in \bar{A}^{n+r-1}, \quad t \in I), \\ y_0 & (a \in A, \quad t = I). \end{cases}$$

Since $F_0 = F$ is defined over $\bar{B}^{n+r}$, it follows from the homotopy extension property [2, p. 501] that F_λ has an extension $F^*_\lambda\colon \bar{B}^{n+r} \to Y$ $(o \leqq \lambda \leqq 1)$ such that $F^*_0 = F$. Clearly we have $F^*_1[(\bar{A}^{n+r-1} \times I) \cup (A \times 1)] = y_0$, and $F^*_\lambda(B_1) = y_0$ for each $o \leqq \lambda \leqq 1$. Define a mapping $\phi\colon A \to Y$ by taking $\phi(a) = F^*_1(a, o)$, $(a \in A)$. Since $\phi(\bar{A}^{n+r-1}) = y_0$, ϕ represents an element $\alpha \in K_{n-1}$. It is obvious that $\mu_n(\alpha) \in H_*^n(X_*, \pi_{n+r})$ is the coset which contains the image under ι of the element

$$\gamma^{n+r+1}(F^*_1) = \gamma^{n+r+1}(F) = \xi \in P^{n+r+1}(B_*, \pi_{n+r}).$$

This proves that μ_n is onto.

To consider the kernel of μ_n, suppose $\alpha \in K_{n-1}$ and $\mu_n(\alpha) = 0$. Let $f\colon A \to Y$, $f(\bar{A}^{n+r+1}) = y_0$, be a representative of α and define $F\colon \bar{B}^{n+r} \to Y$ as above. Then the element $\gamma^{n+r+1}(F)$ is regular; and hence there exists a mapping $G\colon \bar{B}^{n+r} \to Y$, $G(B_0) = y_0$, such that $\gamma^{n+r+1}(G) = \gamma^{n+r+1}(F)$. Define a mapping $\Phi\colon B^{n+r} \to Y$ by taking

$$\Phi(x, e_1, e_2, \cdots, e_r, t) = \begin{cases} F(x, 2e_1, e_2, \cdots, e_r, t) & (o \leqq e_1 \leqq \tfrac{1}{2}), \\ G(x, 2 - 2e_1, e_2, \cdots, e_r, t), & (\tfrac{1}{2} \leqq e_1 \leqq 1), \end{cases}$$

for each point $(x, e_1, e_2, \cdots, e_r, t) \in \bar{B}^{n+r} \subset B$, where $x \in X$, $(e_1, e_2, \cdots, e_r) \in E^r$, $t \in I$. Then $\Phi(B_1) = y_0$ and

$$\gamma^{n+r+1}(\Phi) = \gamma^{n+r+1}(F) - \gamma^{n+r+1}(G) = 0.$$

It follows from the first extension theorem of S. Eilenberg [3, p. 239] that the partial mapping $\Phi \mid \bar{B}^{n+r-1}$ has an extension $\Phi^*\colon \bar{B}^{n+r+1} \to Y$. Define a homotopy

$$\Psi_\lambda\colon (\bar{A}^{n+r} \times I) \cup (A \times 1) \to Y \qquad (o \leqq \lambda \leqq 1)$$

by taking

$$\Psi_\lambda(a, t) = \begin{cases} \Phi^*(a, \lambda + t - \lambda t) & (a \in \bar{A}^{n+r}, \quad t \in I), \\ y_0 & (a \in A, \quad t = 1). \end{cases}$$

Since $\Psi_0 = \Phi^*$ is defined over $\bar{B}^{n+r+1}$, it follows from the homotopy extension property [2, p. 501] that Ψ_λ has an extension $\Psi^*_\lambda\colon \bar{B}^{n+r+1} \to Y$, $(o \leqq \lambda \leqq 1)$, such that $\Psi^*_0 = \Phi^*$. Define a homotopy $f_t\colon A \to Y$ $(o \leqq t \leqq 1)$ by taking $f_t(a) = \Psi^*_t(a, o)$, $(a \in A, o \leqq t \leqq 1)$. Since $G(A \times o) = y_0$ and $\Psi^*_0 = \Phi^*$, f_0 clearly represents the same element α as f does. Since $\Psi^*_1(\bar{A}^{n+r} \times I) = y_0$, we have $f_1(\bar{A}^{n+r}) = y_0$. Further, it follows from $\Psi^*_t(B_1) = y_0$ that $f_t(A_0) = y_0$ for each $o \leqq t \leqq 1$. Hence α is n-trivial.

Conversely, suppose $\alpha \in K_n$ be represented by a mapping $f\colon A \to Y$, $f(\bar{A}^{n+r}) = y_0$. Define $F\colon \bar{B}^{n+r+1} \to Y$ by taking

$$F(a, t) = \begin{cases} f(a) & (a \in A, \quad t = 0), \\ y_0 & (a \in \bar{A}^{n+r}, \quad t \in I), \\ y_0 & (a \in A, \quad t = 1); \end{cases}$$

then clearly $\gamma^{n+r+1}(F) = 0$ and hence $\mu_n(\alpha) = 0$. Therefore, the kernel of μ_n is the n-trivial subgroup K_n. This completes the proof of (4. 1).

(4. 2) *Each group of the sequence*

$$K_0 \supset K_1 \supset \cdots \supset K_n \supset \cdots \supset K_m = 1$$

is a normal subgroup of the preceding and each of the quotient groups K_{n-1}/K_n, $(n = 1, 2, \cdots, m)$, *is abelian and isomorphic with* $H_{\bullet}^n(X_{\bullet}, \pi_{n+r})$. *Hence, the subgroups* K_n $(n = o, 1, \cdots, m)$ *are solvable groups* [7, p. 15].

Proof. Since K_n is the kernel of μ_n, it is a normal subgroup of K_{n-1}. Since μ_n is onto, it follows from Noether's theorem that $K_{n-1}/K_n \approx H_{\bullet}^n(X_{\bullet}, \pi_{n+r})$. Since K_{n-1}/K_n is abelian, the commutator subgroup of K_{n-1} is contained in K_n. Hence it can be easily seen that K_{n-1} is solvable. Q. E. D.

5. The tubular homomorphism. Let $\alpha \in \pi_r(\Omega_0)$ be an arbitrary element, and $f\colon A \to Y$, $f(A_0) = y_0$, be a representative of α. Choose a point $x_0 \in X$. Define a mapping $\phi\colon E^r \to Y$ by taking $\phi(e) = f(x_0, e)$, $(e \in E^r)$. Since $\phi(\partial E^r) = y_0$, ϕ represents an element $\xi \in \pi_r(Y)$. Since X is connected, ξ does not depend on the choice of x_0. Further, it is clear that ξ is independent of the choice of the representative f. Hence ξ depends only on α and will be denoted by $\tau_r(\alpha)$.

(5. 1) *The correspondence* $\alpha \to \tau_r(\alpha)$ *is a homomorphism* (*called the tubular homomorphism* $\tau_r\colon \pi_r(\Omega_0) \to \pi_r(Y)$. τ_r *is onto if* x_0 *is vacuous; otherwise,* τ_r *maps* $\pi_r(\Omega_0)$ *into the identity element of* $\pi_r(Y)$. *The kernel* K *of* τ_r *is the o-trivial subgroup* K_0 *of* $\pi_r(\Omega_0)$.

Proof. That τ_r is a homomorphism is trivial. If X_0 is non-empty, we may choose $x_0 \varepsilon X_0$; hence $\phi(E^r) = y_0$ and $\tau_r(\alpha) = 1$ for each element $\alpha \varepsilon \pi_r(\Omega_0)$. Suppose X_0 be empty. Let an arbitrary element $\xi \varepsilon \pi_r(Y)$ be represented by $\phi\colon E^r \to Y$ with $\phi(\partial E^r) = y_0$. Define a mapping $f\colon A \to Y$ by taking $f(x, e) = \phi(e)$, $(x \varepsilon X, e \varepsilon E^r)$. Since X_0 is void, $A_0 = X \times \partial E^r$; hence we have $f(A_0) = y_0$. f represents an element $\alpha \varepsilon \pi_r(\Omega_0)$. Clearly $\tau_r(\alpha) = \xi$. Therefore, τ_r is onto.

To consider the kernel of τ_r, let us assume $\alpha \varepsilon \pi_r(\Omega_0)$ with $\tau_r(\alpha) = 1$. Let $f\colon A \to Y$, $f(A_0) = y_0$, be a representative of α. For each r-cell $s_i = x_i \times E^r$, where x_i is a vertex of $X_\bullet$, the partial mapping $f \mid s_i$ represents the element $\tau_r(\alpha) = 1 \varepsilon \pi_r(Y)$. Hence, it follows from the homotopy extension property [2, p. 501] that there exists a homotopy $f_t\colon A \to Y$, $(o \leqq t \leqq 1)$, such that $f_0 = f$, $f_1(\bar{A}^r) = y_0$, and $f_t(A_0) = y_0$ for each $o \leqq t \leqq 1$. Therefore, $\alpha \varepsilon K_0$.

Conversely, suppose $\alpha \varepsilon K_0$ and $f\colon A \to Y$, $f(\bar{A}^r) = y_0$, be a representative of α. If we choose x_0 to be a vertex of X, we see that $\tau_r(\alpha) = 1$. Hence the kernel of τ_r is $K = K_0$. This completes the proof.

Hereafter, the kernel $K = K_0$ of the tubular homomorphism τ_r will be called the *principal subgroup* of $\pi_r(\Omega)$. The following statements are immediate consequences of (5.1).

(5.2) *If X_0 is non-empty, then $\pi_1(\Omega_0) = K_0$ is a solvable group.*

(5.3) *If X_0 is empty, then $K = K_0$ is a normal subgroup of $\pi_r(\Omega_0)$ and $\pi_r(\Omega_0)/K \approx \pi_r(Y)$.*

In the remainder of this section, we shall assume that X_0 be empty.

An element $\alpha \varepsilon \pi_r(\Omega_0)$ is said to be *tubular* if it can be represented by a mapping $f\colon A \to Y$, $f(A_0) = y_0$, such that $f(x_1, e) = f(x_2, e)$ for arbitrary points $x_1, x_2 \varepsilon X$, $e \varepsilon E^r$. The tubular elements of $\pi_r(\Omega_0)$ clearly form a subgroup T, called the *tubular subgroup* of $\pi_r(\Omega_0)$.

(5.4) *The tubular homomorphism τ_r maps the tubular subgroup T of $\pi_r(\Omega_0)$ isomorphically onto $\pi_r(Y)$.*

Proof. That τ_r maps T onto $\pi_r(Y)$ has been proved in the first paragraph of the proof of (5.1). To prove that τ_r maps T isomorphically, let us suppose that $\alpha \varepsilon T$ and $\tau_r(\alpha) = 1$. Let $f\colon A \to Y$ be a representative of α as described above. Choose a point $x_0 \varepsilon X$ and define a mapping $\phi\colon E^r \to Y$ by taking $\phi(e) = f(x_0, e)$ for each $e \varepsilon E^r$. Since ϕ represents the identity element of $\pi_r(Y)$, there exists a homotopy $\phi_t\colon E_r \to Y$, $(o \leqq t \leqq 1)$, such that $\phi_0 = \phi$, $\phi_1(E^r) = y_0$, and $\phi_t(\partial E^r) = y_0$ for each $o \leqq t \leqq 1$. Define a homotopy $f_t\colon A \to Y$, $(o \leqq t \leqq 1)$, by taking $f_t(x, e) = \phi_t(e)$ for each $x \varepsilon X$, $e \varepsilon E^r$, and each $o \leqq t \leqq 1$. Then we have $f_0 = f$, $f_1(A) = y_0$, and $f_t(A_0) = y_t$ for each $o \leqq t \leqq 1$. Hence, $\alpha = 1$ and the proof is complete.

(5.5) *$\pi_r(\Omega_0)$ is the direct product of its tubular subgroup T and its principal subgroup K if and only if T is a normal subgroup of $\pi_r(\Omega)$.*

Proof. The necessity is trivial. To prove the sufficiency, let us suppose T to be a normal subgroup of $\pi_r(\Omega_0)$. Then both T and K are normal subgroups of $\pi_r(\Omega_0)$. Let $\alpha \varepsilon \pi_r(\Omega_0)$ be an arbitrary element, and let $\xi = \tau_r(\alpha) \varepsilon \pi_r(Y)$. Since τ_r maps T isomorphically onto $\pi_r(Y)$, there exists a unique element $\delta \varepsilon T$ such that $\tau_r(\delta) = \xi$. Then the element $\beta = \delta^{-1}\alpha$ is contained in the principal subgroup K; and we have $\alpha = \delta\beta$, i. e. $\pi_r(\Omega_0) = TK$. Further, since τ_r maps T isomorphically and K is the kernel of τ_r, we have $T \cap K = 1$. Hence, $\pi_r(\Omega_0)$ is the direct product of T and K, [7, p. 16]. This completes the proof.

Since $\pi_r(\Omega_0)$ is abelian when $r > 1$, we have the following corollary of (5.5).

(5.6) *For each $r > 1$, $\pi_r(\Omega_0)$ is the direct sum of T and K.*

Thus we have proved all the results sketched in the introduction.

6. Specializations. In the present section, we shall deduce some interesting particular cases of our general results concerning the structure of the principal subgroup K of $\pi_r(\Omega_0)$.

(6.1) *The principal subgroup K of $\pi_r(\Omega_0)$ is isomorphic with the cohomology group $H^n(X_*, \pi_{n+r})$ if the following conditions are satisfied:*

(i) $H^{p-1}(X_*, \pi_{p+r}) = 0$ *for each* $1 < p < n$,

(ii) $H^{p+1}(X_*, \pi_{p+r}) = 0$ *for each* $n < p < m$,

(iii) $H^p(X_*, \pi_{p+r}) = 0$ *except* $p = 0$ *and* $p = n$.

Proof. According to (3.3), the condition (i) implies that $H_*^n(X_*, \pi_{n+r})$ is isomorphic with $P^n(X_*, \pi_{n+r})$. According to (3.4), the condition (ii) implies that $P^n(X_*, \pi_{n+r})$ coincides with $H^n(X_*, \pi_{n+r})$. Hence we obtain the isomorphism

$$H_*^n(X_*, \pi_{n+r}) \approx H^n(X_*, \pi_{n+r}).$$

It follows from (iii) that $H_*^p(X_*, \pi_{n+r}) = 0$ for each $p > 0$ and $p \neq n$. Hence K is isomorphic with the cohomology group $H^n(X_*, \pi_{n+r})$. Q. E. D.

As a corollary of (6. 1), we give the following theorem.

(6. 2) *If* $\pi_p(Y) = 0$ *for each* $r < p < m + r$, *then the principal subgroup* K *of* $\pi_r(\Omega_0)$ *is isomorphic with the cohomology group* $H^m(X_\bullet, \pi_{m+r})$.

(6. 3) *If the integral homology groups* $H_p(X_\bullet) = 0$ *for each* $o < p < m$, *then the principal subgroup* K *of* $\pi_r(\Omega_0)$ *is isomorphic with the cohomology group* $H^m(X_\bullet, \pi_{m+r})$.

Proof. Since $H_p(X_\bullet) = 0$ for each $o < p < m$, it follows from the first duality theorem [6, p. 117] that the integral cohomology groups $H^p(X_\bullet) = 0$, $(o < p < m)$, and the m-dimensional cotorsion group $T^m(X_\bullet) = 0$. Then it follows from a theorem of Eilenberg and MacLane [6, p. 347] that $H^p(X_\bullet, G) = 0$, $(o < p < m)$, for each discrete coefficient group G. Hence the conditions (i)-(iii) of (6. 1) are satisfied with $n = m$. This completes the proof.

In particular, if we put $X = S^m$, the m-sphere, and X_0 either vacuous or a single point, then (5. 5), (5. 6), (6.3) include the results of M. Abe [1] and the author [5].

INSTITUTE OF MATHEMATICS, ACADEMIA SINICA,
NANKING, CHINA.

BIBLIOGRAPHY.

[1] Abe, M., "Ueber die stetigen Abbildungen der n-Sphäre in einen metrischen Raum," *Japanese Journal of Mathematics*, vol. 16 (1939), pp. 169-176.
[2] P. Alexandroff und H. Hopf, *Topologie* I (1935).
[3] S. Eilenberg, "Cohomology and continuous mappings," *Annals of Mathematics*, vol. 41 (1940), pp. 231-251.
[4] R. H. Fox, "On topologies for function spaces," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 429-432.
[5] S. T. Hu, "On spherical mappings in a metric space," *Annals of Mathematics*, vol. 48 (1947), pp. 717-734.
[6] S. Lefschetz, *Algebraic topology*, American Mathematical Society Colloquium Publications, New York (1942).
[7] L. Pontrjagin, *Topological groups*, Princeton University Press (1939).
[8] N. E. Steenrod, "Products of cocycles and extension of mappings," *Annals of Mathematics*, vol. 48 (1947), pp. 290-320.

A PRIORI LAPLACE TRANSFORMATIONS OF LINEAR DIFFERENTIAL EQUATIONS.*

By Aurel Wintner.

1. The results of the following considerations are in the direction initiated in [9] and applied in [10]. They deal with the solution of certain differential equations

$$x'' + f(t)x = 0 \tag{1}$$

(on an unbounded interval

$$t^* < t < \infty, \tag{2}$$

with an unspecified t^*) by means of definite integrals of specific type.

The underlying idea is similar to, but the assumptions and the results are substantially distinct from, those in [9]. Correspondingly, [9] will not be needed.

2. If $f(t)$ is a real-valued, continuous function on an unspecified half-line (2), the differential equation (1) is called oscillatory or non-oscillatory according as every or no real-valued solution $x(t) \not\equiv 0$ changes sign an infinity of times as $t \to \infty$. In view of Sturm's separation theorem, (1) must be non-oscillatory if it is not oscillatory.

As observed by Kneser [3], the differential equation (1) cannot be oscillatory if

$$-\infty \leqq \limsup_{t\to\infty} t^2 f(t) < \tfrac{1}{4}, \tag{3}$$

whereas it must be oscillatory if

$$\tfrac{1}{4} < \liminf_{t\to\infty} t^2 f(t) \leqq \infty. \tag{3 bis}$$

The proof of these criteria follows immediately, if (1) is "compared," in Sturm's sense, with the trivial differential equation

$$x'' + ct^{-2}x = 0. \tag{4}$$

In fact, substitution of

$$x(t) = t^\lambda$$

---

* Received September 17, 1948.

into (4) gives

$$2\lambda = 1 \pm (1 - 4c)^{\frac{1}{2}},$$

and the last two formula lines show that (4) is oscillatory when $c > \frac{1}{4}$ and non-oscillatory when $c < \frac{1}{4}$.

In view of Sturm's comparison theorem, "logarithmic" refinements of the criteria (3), (3 bis) are corollaries of the asymptotic formulae of Hartman [1] for the solutions of (1) when $f(t)$ is an arbitrary logarithmico-exponential function. A criterion of quite another, though still explicit, type was given in [12].

None of these explicit criteria supplies a condition which is necessary *and* sufficient in order that (1) be oscillatory (or non-oscillatory). A condition which is both necessary and sufficient (but, instead of being "explicit," is of the "Lebesgue-Toeplitz" type) was given in [11].

3. The content of the theorem to be proved below can be illustrated by the following corollary of it:

If $\sum_{n=1}^{\infty} c_n z^n$ *converges everywhere, and if*

$$4c_1 < 1 \tag{5}$$

and

$$c_n \geqq 0 \qquad (n = 1, 2, \cdots), \tag{6}$$

then the linear differential equation

$$x'' + x \sum_{n=1}^{\infty} (-1)^{n+1} c_n / t^{2n} = 0 \tag{7}$$

has, for large positive t, two linearly independent solutions representable in the form

$$x(t) = \exp - \left(\int_0^{\infty} \right) (e^{-i(t-t_0)s} - 1)/s \, d\phi(s), \tag{8}$$

with

$$(-i)^n d^n \log x(t)/dt^n = - \left(\int_0^{\infty} \right) s^n e^{-i(t-t_0)s} \, d\phi(s), \qquad (n > 0), \tag{8 bis}$$

where $\phi(s)$ *is a non-decreasing function, the convergence of* (8), (8 bis) *on the half-line* $t_0 < t < \infty$ *is part of the statement, and* t_0 *and the monotone function* $\phi(s)$ *depend on the pair of integration constants determining a solution* $x(t)$ *of* (7).

The parentheses of the integral signs in (8), (8 bis) refer to Abelian "summations"; so that the convergence of

$$\Bigl(\int_0^\infty\Bigr) g(s)\,ds$$

means that the integral

$$G(\epsilon) = \int_0^\infty e^{-\epsilon s} g(s)\,ds$$

converges for every $\epsilon > 0$ and tends to a finite limit as $\epsilon \rightarrow 0$.

If t is replaced by ti, then (7) goes over into the case

$$f(t) = \sum_{n=1}^{\infty} c_n/t^{2n}$$

of (1). In view of (5), this $f(t)$ satisfies (3) and, by virtue of (6), is such that the inequality

$$(9_k) \qquad (-1)^k D^k f(t) \geqq 0, \text{ where } k = 0, 1, \cdots, \qquad (D^k = d^k/dt^k),$$

holds on the half-line $0 < t < \infty$. Accordingly, (1) is non-oscillatory and its coefficient function is completely monotone.

4. If this situation is compared with the substitution, $t \rightarrow ti$, applied to (7), it is seen that the last italicized statement can be obtained by replacing t by $t + i\epsilon$, and then letting ϵ (> 0) tend to 0, in the assertion, (12), of the following theorem:

On some half-line (2), *let* $f(t)$ *be a function possessing arbitrarily high derivatives which satisfy the inequalities* (9). *Suppose further that* $f(t)$ *is such as to render the differential equation* (1) *non-oscillatory* (*e. g., that* (3) *is satisfied by* f). *Then there exists on* (2) *a* $t = T$ *having the following property*: *There belongs to every* t_0 *satisfying*

$$(10) \qquad t_0 > T \qquad (T > t^*)$$

a solution of (1) *representable on the half-line*

$$(11) \qquad t_0 < t < \infty$$

in the form

$$(12) \qquad x(t) = \exp - \int_0^\infty (e^{-(t-t_0)s} - 1)/s \, d\phi(s),$$

where $\phi(s)$ *depends on* t_0, *is a non-decreasing function of s, i. e.,*

$$d\phi(s) \geqq 0 \text{ for } 0 \leqq s < \infty, \tag{13}$$

and is such that

$$\int_0^\infty e^{-\epsilon s} d\phi(s) < \infty \text{ if } \epsilon > 0 \tag{14}$$

but

$$\int_0^\infty d\phi(s) = \infty; \tag{15}$$

so that, by (12),

$$x(t) \neq 0 \text{ on } (11) \tag{14 bis}$$

but

$$x(t_0) = 0. \tag{15 bis}$$

The proof will depend on the replacement of (1) by

$$y' + y^2 + f(t) = 0, \tag{16}$$

the Riccati resolvent of (1). It is satisfied by the logarithmic derivative,

$$y(t) = x'(t)/x(t), \tag{17}$$

of every solution $x(t) \not\equiv 0$ of (1). If t_0 is large enough, the denominator of (17) does not vanish on (11), since (1) is supposed to be non-oscillatory.

5. Let $x(t)$ be any real-valued, non-trivial ($\not\equiv 0$) solution of (1). Since $x(t)$ can be replaced by $-x(t)$, and since (1) is non-oscillatory, it can be assumed that

$$x(t) > 0 \tag{18}$$

on (11), provided the t_0 defining (11) is chosen large enough. It will be assumed that t_0 is fixed in accordance with this proviso, and that t is restricted to the corresponding half-line (11).

According to (9_0), the function $f(t)$ is non-negative. It follows therefore from (1) and (18) that

$$x''(t) \leqq 0. \tag{18 bis}$$

On the other hand, it is clear, for reasons of convexity, that (18) *and* (18 bis) cannot hold on the entire half-line (11) unless $x(t)$ is non-decreasing, hence $x'(t) \geqq 0$, on (11). It follows therefore from (17) and (18) that $y(t) \geqq 0$.

This proves that, if $D^n = d^n/dt^n$, the case $n = 0$ of the inequality

$$(19_n) \qquad (-1)^n D^n y(t) \geqq 0$$

is true on (11). Since $f(t)$ is non-negative, (16) shows that (19_n) holds for $n = 1$ also. In order to prove that (19_n) is true for every n, it will be assumed that

$$(20) \qquad (19_0),\ (19_1), \cdots, (19_k) \text{ are true,}$$

and it will be shown that (20) implies (19_{k+1}) by virtue of (9_k).

First, if the binomial coefficients are denoted by $C = C(k, j)$, it is seen from (16), by k-fold differentiation, that

$$-D^{k+1}y = D^k f + 2\sum_{j=0}^{k-1} C(k-1, j)(D^{j+1}y)(D^{k-1-j}y).$$

It follows therefore from (9_k) that (19_{k+1}) is true if

$$(-1)^k \sum_{j=0}^{k-1} C(k-1, j)(D^{j+1}y)(D^{k-1-j}y) \geqq 0.$$

Since $C(k-1, j) > 0$, it follows that it is sufficient to ascertain the truth of the k inequalities

$$(-1)^k (D^{j+1}y)(D^{k-1-j}y) \geqq 0, \text{ where } 0 \leqq j \leqq k-1.$$

But these k inequalities are implied by the $k+1$ assumptions (20).

6. This proves that (19_n) holds on (11) for every n. In other words, $y(t)$ is completely monotone on (11), i. e., $y(t + t_0)$ is completely monotone on the half-line $0 < t < \infty$. Hence, the Hausdorff-Bernstein theorem supplies the existence of a function $\phi(s)$ satisfying (13) and having the property that

$$y(t + t_0) = \int_0^\infty e^{-st} d\phi(s)$$

holds on the half-line $0 < t < \infty$, i. e.,

$$(12 \text{ bis}) \qquad y(t) = \int_0^\infty e^{-(t-t_0)s} d\phi(s)$$

on the half-line (11). The existence of an $x(t)$ represented on (11) by (12) now follows from (17).

The specification (15) was in no sense involved thus far. In fact, (12) and (13) show that (15) is equivalent to (15 bis), whereas the requirement imposed above on t_0 was nothing like (15 bis), but merely the proviso that t_0 is large enough.

In order to satisfy (15 bis), it is sufficient to combine the latter proviso with the fact that $x(t^0) = 0$ (along with an arbitrary $x'(t^0) \neq 0$) can be assigned as an initial condition for a solution $x(t) \not\equiv 0$ of (1).

7. In order to complete the proof of the last italicized theorem, it is sufficient to apply the following remark:

If a differential equation (1), *in which* $f(t)$ *denotes a real-valued, continuous function for large positive* t, *is non-oscillatory, then there exists a* T *having the property that no* (*real-valued, non-trivial*) *solution* $x(t)$ *of* (1) *has more than one zero on the half-line*

$$T < t < \infty. \tag{10 bis}$$

Suppose that there does not exist such a T. Then (1) has (real-valued, non-trivial) solutions $x_1(t), x_2(t), \cdots$ the n-th of which has the property that $x_n(u_n) = 0$ and $x_n(v_n) = 0$, where $u_n < v_n$, and $u_n \to \infty$ as $n \to \infty$. But Sturm's separation theorem claims that $x_1(t)$ must have a zero on the interval $u_n \leqq t \leqq v_n$. It follows therefore from $u_n \to \infty$ that the zeros of $x_1(t)$ cluster at $t = \infty$. This contradicts the assumption that (1) is non-oscillatory.

APPENDIX.

The considerations of [9] depended on the following elementary fact which, in a slightly different formulation, is due to Kneser [4]:

If $f(t)$ is real-valued, continuous and non-positive on (2), then (1) has a solution satisfying

$$x(t) > 0 \text{ and } x'(t) \leqq 0 \tag{21}$$

on (1).

This fact leads, again in the direction of [9], to a curious result, which can be formulated as follows:

If $f(t)$ *is real-valued, non-positive and continuous for* $0 \leqq t < \infty$, *then* (1) *has a solution representable in the form*

$$x(t) = \text{const.} + \int_0^\infty X(s) \cos ts \, ds \text{ for } 0 < t < \infty, \tag{22}$$

where

$$\text{const.} = \lim_{t\to\infty} x(t) \tag{23}$$

and

$$X(s) > 0. \tag{24}$$

The existence of the limit (23), which is part of the statements, cannot be concluded from the balance of the statements. In fact, (23) can be concluded from (22) only if the integral analogue of the Riemann-Lebesgue lemma is applicable to the Fourier transform occurring in (22). In view of (24), this is the case only if $X(s)$ is absolutely integrable over $0 \leqq s < \infty$. But this is not claimed, since $t = 0$ is excluded in (22).

It may be mentioned that, in (22),

$$\text{const.} = 0 \text{ if and only if } \int_0^\infty tf(t)\,dt = \infty \tag{25}$$

in (1), where $f \leqq 0$ (for a simple proof of the characterization (25) of the vanishing of the limit (23), cf. [2], Appendix; the method applied there leads to a result sharper than what is claimed by the first of the assertions of (25)).

First, if $y(t)$, where $0 \leqq t < \infty$, is any function possessing a continuous second derivative and satisfying

$$y(t) > 0, \qquad y'(t) \leqq 0, \qquad y''(t) \geqq 0 \tag{26}$$

and if, in addition to (26),

$$y(t) \to 0 \text{ as } t \to \infty, \tag{27}$$

then the Fourier cosine transform,

$$\int_0^\infty y(s) \cos ts\, ds, \tag{28}$$

of $y(t)$ is convergent, and represents a non-negative function on the half-line $0 < t < \infty$. This follows, by a partial integration, from the fact that

$$\sum_{n=1}^{\infty} (-1)^{n+1} a_n \geqq 0 \text{ if } 0 \leqq a_n \leqq a_{n+1} \to 0;$$

cf. [5], or [6], p. 378, where (27) is omitted.

Next, if $x(t)$ is a solution of (1) supplied, if $f \leqq 0$, by Kneser's theorem, then, since (21) holds on the half-line $0 \leqq t < \infty$, the function $x(t)$ is positive and tends non-increasingly to a limit $x(\infty)$. Furthermore, $x''(t) \geqq 0$, by (1), where $f(t) \leqq 0$ and $x(t) > 0$. Consequently, (27) and all three conditions (26) are satisfied by

$$y(t) = x(t) - x(\infty), \tag{29}$$

where $0 \leqq t < \infty$.

The last italicized theorem follows by substituting (29) into (28) and applying a Fourier inversion. In fact, (27) and (26) together are more than sufficient for the applicability of a standard criterion for the legitimacy of Fourier's inversion (if $t \neq 0$); cf. [7].

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman, "On the linear logarithmico-exponential differential equation of the second order," *American Journal of Mathematics*, vol. 70 (1948), pp. 765-779.

[2] ——— and A. Wintner, "On the Laplace-Fourier transcendents," *ibid.*, vol. 71 (1949), pp. 367-372.

[3] A. Kneser, "Untersuchungen über die reellen Nullstellen der Integrale linearer Differentialgleichungssysteme," *Mathematische Annalen*, vol. 42 (1893), pp. 409-435.

[4] ———, "Untersuchung und asymptotische Darstellung der Integrale gewisser Differentialgleichungen bei grossen reellen Werthen des Arguments, I," *Journal für die reine und angewandte Mathematik*, vol. 116 (1896), pp. 178-212.

[5] A. Lindhagen, *Studier öfver Gamma-Funktionen och nagra beslagtade Transcendenter*, Uppsala, 1887; quoted from [8], p. 184.

[6] G. Pólya, "Ueber die Nullstellen gewisser ganzer Funktionen," *Mathematische Zeitschrift*, vol. 2 (1918), pp. 352-383.

[7] A. Pringsheim, "Ueber neue Gültigkeitsbedingungen für die Fouriersche Integralformal," *Mathematische Annalen*, vol. 68 (1910), pp. 367-408; Nachtrag, *ibid.*, vol. 71 (1912), pp. 289-298.

[8] J. F. Steffensen, "Bounds of certain trigonometrical integrals," *Tenth Scandinavian Congress of Mathematicians*, Copenhagen, 1947, pp. 181-186.

[9] A. Wintner, "On the Laplace-Fourier transcendents occurring in mathematical physics," *American Journal of Mathematics*, vol. 69 (1947), pp. 87-98.

[10] ———, "On the normalization of characteristic differentials in continuous spectra," *Physical Review*, vol. 72 (1947), pp. 516-517.

[11] ———, "A norm criterion for non-oscillatory differential equations," *Quarterly of Applied Mathematics*, vol. 6 (1948), pp. 183-185.

[12] ———, "A criterion of oscillatory stability," *ibid.*, vol. 7 (1949), pp. 115-117.

ON ALMOST FREE LINEAR MOTIONS.*

By Aurel Wintner.

1. For large positive t, say for

$$t_0 < t < \infty, \tag{1}$$

let $f(t)$ be a continuous function having the following properties: The improper integral

$$\int^{\infty} f(t)\,dt = \lim_{T\to\infty} \int^{T} f(t)\,dt \text{ is convergent} \tag{2}$$

(it need not converge absolutely) and, when considered as a function of the lower limit of integration, is such as to make

$$\int^{\infty} \left(\max_{t \leqq v < \infty} \left| \int_{v}^{\infty} f(u)\,du \right| \right) dt < \infty. \tag{3}$$

It will be proved that these assumptions are sufficient in order to make $f(t)$ asymptotically negligible in the homogeneous, linear differential equation

$$x'' + f(t)x = 0 \tag{4}$$

(which is the disturbed form of the trivial differential equation

$$y'' = 0, \tag{5}$$

that of the free motion). By this is meant that, if $x(t)$ is a non-trivial ($\not\equiv 0$) solution of (4), there exists a non-trivial solution, $y(t) = c_1 t + c_2 \not\equiv 0$, of (5) satisfying

$$x(t) \sim y(t) \text{ as } t \to \infty; \tag{6}$$

and that, if $(c_1, c_2) \neq (0, 0)$ is a non-trivial pair of integration constants of $y(t)$, there exists a solution $x(t)$ of (4) satisfying (6).

2. A weaker result is contained in Bôcher's extension of the Fuchsian theory of "regular" singular points to real domain; cf. [1]. In fact, if Bôcher's independent variable is replaced by e^{-t}, where t is the present

---

* Received October 15, 1948.

independent variable, his result for the case of a multiple elementary divisor (*loc. cit.*, § 4) leads to the following criterion: If $f(t)$ is continuous on (1) and if

$$\int^{\infty} t\,|\,f(t)\,|\,dt < \infty, \tag{7}$$

then (4) has a solution satisfying

$$x(t) \sim t \tag{8}$$

and another, linearly independent, solution satisfying

$$x(t) \to 1, \tag{9}$$

as $t \to \infty$.

Clearly, the existence of such a pair of solutions is equivalent to the one-to-one correspondence (6) between the solutions of (4) and (5). But (7) is more severe than (3) with (2). For, on the one hand, (7) implies, whilst (2) and (3) do not imply, that

$$\int^{\infty} |\,f(t)\,|\,dt < \infty,$$

and, on the other hand, (7) is equivalent to

$$\int^{\infty} \Big(\int_{v}^{\infty} |\,f(u)\,|\,du \Big) dv < \infty$$

(Fubini), which requires much more than (3).

3. According to (2),

$$F(t) = \int_{t}^{\infty} f(s)\,ds \tag{10}$$

defines on (1) a function. Since this function is continuous and tends to 0 as $t \to \infty$,

$$G(t) = \max_{t \leqq s < \infty} |\,F(s)\,| \tag{11}$$

defines on (1) a function. Finally, (3) means that

$$H(t) = \int_{t}^{\infty} G(s)\,ds \tag{12}$$

defines on (1) a function. Hence, the assertion, to be proved, is equivalent to the statement that, if $f(t)$ is continuous and such that (10), (11), (12) define functions on (1), then (4) has two solutions satisfying (8), (9) respectively.

It will be sufficient to prove the existence of the solution satisfying (9). In fact, if $f(t)$ is any continuous function on (1) and if $x(t)$ is any function which does not vanish for $T \leqq t < \infty$, where $T > t_0$, two differentiations show that the product

$$x(t) \int_T^t \{x(s)\}^{-2}\, ds \qquad (T \leqq t < \infty) \tag{13}$$

is a solution of (4) whenever $x(t)$ is. But if the solution $x(t)$ satisfies (9), then the solution (13) is of the form

$$\{1 + o(1)\} \int_T^t \{1 + o(1)\}^{-2}\, ds = \{1 + o(1)\}\{t + o(t)\} = t + o(t),$$

as required by (8).

Accordingly, it is more than sufficient to prove the existence of a solution satisfying

$$x(t) \rightarrow 1 \ \textit{and}\ x'(t) \rightarrow 0 \qquad (t \rightarrow \infty). \tag{14}$$

4. To this end, it will first be shown that the recursion formula

$$z_{n+1}'(t) = F(t) z_n(t) + \int_t^\infty F(s) z_n'(s)\, ds, \tag{15}$$

where

$$z_1'(t) = F(t), \tag{16}$$

and the "initial" conditions

$$z_n(\infty) = 0, \tag{17}$$

where $n = 1, 2, \cdots$, define on (1) a sequence of functions, $z_1(t), z_2(t), \cdots$; and that

$$w_n(t) = \int_t^\infty | z_n'(s) |\, ds \tag{18}$$

exists ($< \infty$) for $n = 1, 2, \cdots$ and for every t contained in (1).

The three functions (10), (11), (12) are supposed to exist. Hence,

$$z_1(t) = -\int_t^\infty F(s)\,ds$$

defines a function satisfying (16) and the case $n = 1$ of (17), and (16) shows that the case $n = 1$ of (18) defines a function, $w_1(t)$. In order to apply an induction, suppose that $z_n(t)$ and $w_n(t)$ exist for a fixed n.

According to (17) and (18),

$$|z_n(t)| \leqq w_n(t).$$

Hence, from (11),

$$\int_t^\infty |F(s)z_n(s)|\,ds \leqq G(t)w_n(t),$$

since, according to (18), the function $w_n(t)$ is non-increasing. On the other hand, by (11) and (18),

$$\int_t^\infty \left(\int_u^\infty |F(s)z_n'(s)|\,ds\right)du \leqq \int_t^\infty G(u)w_n(u)\,du,$$

where, according to (12) and the monotony of $w_n(t)$,

$$\int_t^\infty G(u)w_n(u)\,du \leqq H(t)w_n(t).$$

If the last three formula lines are compared with (15), it is seen that

$$\int_t^\infty |z_{n+1}'(s)|\,ds \leqq \{G(t) + H(t)\}w_n(t).$$

In view of (18), this proves that the function $w_{n+1}(t)$ exists. It also follows that the function

$$z_{n+1}(t) = -\int_t^\infty z_n'(s)\,ds \qquad (z_{n+1}(\infty) = 0)$$

exists. This completes the induction.

In view of (18), the last inequality implies that

$$\int_t^\infty |z_{n+1}'(s)|\,ds \leqq \{G(t) + H(t)\}^n w_1(t).$$

On the other hand, (10), (11), (12) show that, if t_0 is large enough, both

$$G(t) < \tfrac{1}{4} \text{ and } H(t) < \tfrac{1}{4}$$

hold for every t on (1). Finally, from (18) and (16),

$$w_1(t) \leqq \int_t^\infty |F(s)|\, ds \to 0 \text{ as } t \to \infty,$$

by (11) and (12).

5. The last three formula lines imply that

$$\int_t^\infty |z_n'(s)|\, ds < \epsilon(t)/2^n \tag{19}$$

where $\epsilon(t)$ is a positive function satisfying $\epsilon(t) \to 0$ as $t \to \infty$. Moreover,

$$|z_n(t)| < \epsilon(t)/2^n, \tag{20}$$

by (19) and (17).

In terms of the functions $z_n(t)$, define the functions $x_n(t)$ by the recursion formula

$$x_n(t) = z_n(t) + x_{n-1}(t), \tag{21}$$

where

$$x_0(t) \equiv 1. \tag{22}$$

According to (21) and (15),

$$x_n'(t) = F(t)x_{n-1}(t) + \int_t^\infty F(s)x_{n-1}'(s)\, ds, \tag{23}$$

provided that $n > 1$. However, since (22) and the case $n = 1$ of (21) lead to $x_1'(t) = z_1'(t)$, it is seen from (15) and (22) that (23) holds for $n = 1$ also. Finally, from (17) and (21),

$$x_n(\infty) = 1. \tag{24}$$

Clearly, (22), (23) and (24) represent a direct definition of the functions $x_n(t)$. Their indirect definition is

$$x_n(t) = 1 + \sum_{m=1}^{n} z_m(t), \tag{25}$$

by (21) and (22). Since $\sum_{m=1}^{n}\epsilon(t)/2^m < \epsilon(t)$, the latter definition and (19), (20) show that

$$(26) \qquad |1-x_n(t)| < \epsilon(t) \text{ and } \int_t^\infty |x_n'(s)|\,ds < \epsilon(t),$$

and (25) shows that $x_n(t)$ tends, as $n \to \infty$, to a limit, say $x(t)$, satisfying

$$(27) \qquad |x(t)-x_n(t)| \leqq \epsilon(t)/2^n \text{ and } \int_t^\infty |d\{x(t)-x_n(t)\}| \leqq \epsilon(t)/2^n.$$

Since every $x_n(t)$, being differentiable, is continuous, the first of the inequalities (27) implies that $x(t)$ is continuous; and that

$$(28) \qquad x(\infty) = 1,$$

by the first of the relations (26), where $\epsilon(t) \to 0$ as $t \to 0$.

6. In order to carry out the limit process, $n \to \infty$, in (23), a sufficient estimate of the first derivatives of the functions (21) remains to be ascertained.

According to (15) and (11),

$$|z_{n+1}'(t)| \leqq G(t)|z_n(t)| + G(t)\int_t^\infty |z_n'(s)|\,ds.$$

Hence, if n is replaced by $n-1$, it follows from (19) and (10) that

$$(29) \qquad |z_n'(t)| \leqq 4G(t)\epsilon(t)/2^n.$$

This estimate is sufficient to assure the legitimacy of the limit process in question. In fact, $x(t)$ was defined as the limit of $x_n(t)$ as $n \to \infty$. This means, by (25), that

$$(30) \qquad x(t) = 1 + \sum_{n=1}^{\infty} z_n(t).$$

But (29) implies that the derived series of (30) is uniformly convergent. Hence, the function $x(t)$ has a derivative,

$$(31) \qquad x'(t) = \sum_{n=1}^{\infty} z_n'(t).$$

In addition, $x'(t)$ is continuous, by (31) and (29), and since every $z_n'(t)$ is continuous. Moreover, since $G(t) \to 0$ and $\epsilon(t) \to 0$ as $t \to \infty$, it is seen from (31) and (29) that

$$x'(\infty) = 0. \tag{32}$$

It is now clear from the existence ($\neq \infty$) of all three functions (10), (11), (12) that the limit process, $n \to \infty$, leads from (23) to

$$x'(t) = F(t)x(t) + \int_t^\infty F(s)x'(s)\,ds.$$

On the other hand, since (10) implies that $F(t) \to 0$ as $t \to \infty$, it implies, by (28), that $F(t)x(t) \to 0$. Hence, it is seen from (10) that a partial integration gives

$$\int_t^\infty F(s)x'(s)\,ds = -F(t)x(t) + \int_t^\infty f(s)x(s)\,ds.$$

According to the last two formula lines, $x'(t) = \int_t^\infty f(s)x(s)\,ds$ is an identity. Since $f(t)$ and $x(t)$ are continuous, the function on the right of this identity is differentiable. Hence, the same is true of the function on the left. Accordingly, $x''(t) = -f(t)x(t)$. This means that $x(t)$ is a solution of (4). On the other hand, (24) and (32) mean that $x(t)$ satisfies (14). Hence, the proof is complete.

Appendix.

It may be mentioned that, if $f(t)$ is real-valued and does not change its sign for large t, then the pair of conditions (2), (3), which is always sufficient, becomes necessary as well, in order that some solution of (4) should satisfy (9). In fact, the calculation made between (9) and (10) not only shows that (7) always implies (2) and (3), but it also shows that (2) and (3) imply (7) if it is assumed that one and the same of the inequalities

$$\pm f(t) \geqq 0 \tag{33}$$

holds from a certain t onward. Accordingly, all that is claimed is that, in this particular case, (7) is necessary for the existence of a solution satisfying (9). But the truth of this remark is known; it goes back to Weyl [2], § 1. A simple proof proceeds as follows:

Suppose that a continuous $f(t)$ satisfying the alternative (33) fails to satisfy (7). Then

$$\pm \int^{t} sf(s)\,ds \to \infty \qquad (t \to \infty). \tag{34}$$

On the other hand, if (4) had a solution satisfying (9), then, since (4), (9) and (33) imply that

$$-sx''(s) = \pm \mid sx(s)f(s) \mid \tag{35}$$

holds for large s, it would follow from (34) and (9) that $\left| \int^{t} sx''(s)\,ds \right| \to \infty$ as $t \to \infty$. Since a partial integration gives

$$\int^{t} sx''(s)\,ds = \text{const.} + tx'(t) - \int^{t} x'(s)\,ds,$$

and since (9) implies the convergence of the integral $\int^{\infty} x'(t)\,dt$, it follows that $\mid tx'(t) \mid \to \infty$. Hence, if $x'(t)$ does not change sign from a certain t onward, a quadrature shows that $\mid x(t) \mid$ tends, as $t \to \infty$, to ∞ (stronger than any fixed multiple of $\int dt/t = \log t$). But this contradicts (9). Finally, $x'(t)$ cannot change sign from a certain t onward, since $x'(t)$ is ultimately monotone. In fact, (35) shows that $x''(t)$ does not change sign (from a certain t onward).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] M. Bôcher, "On regular singular points of linear differential equations of the second order whose coefficients are not necessarily analytic," *Transactions of the American Mathematical Society*, vol. 1 (1900), pp. 40-52.

[2] H. Weyl, "Ueber gewöhnliche lineare Differentialgleichungen mit singulären Stellen und ihre Eigenfunktionen," *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen*, 1909, pp. 37-63.

ON THE SMALLNESS OF ISOLATED EIGENFUNCTIONS.*

By AUREL WINTNER.

---

1. In the linear differential equation

$$x'' + (f(t) + \lambda)x = 0, \tag{1}$$

where $0 \leqq t < \infty$ and $-\infty < \lambda < \infty$, let $f(t)$ be a real-valued, continuous function satisfying

$$-\infty \leqq \limsup_{t\to\infty} f(t) < \infty. \tag{2}$$

For every value of the parameter λ, consider only real-valued solutions

$$x = x(t) = x_\lambda(t; c_1, c_2); \qquad c_1 = x(0), c_2 = x'(0).$$

If

$$x'' + (f(t) + \lambda_0)x = 0 \tag{3}$$

has a non-trivial solution of class (L^2), that is, a solution $x = y(t)$ satisfying

$$0 < \int_0^\infty y^2(t)\,dt < \infty, \tag{4}$$

then $y(t)$ is called an eigenfunction (belonging to the $\lambda = \lambda_0$ occurring in (3) and to that boundary condition, say

$$x(0)\cos\phi + x'(0)\sin\phi = 0, \tag{5}$$

which is satisfied by $x = y(t)$ at $t = 0$).

According to Weyl [3], p. 238, the assumption (2) precludes the existence of a λ_0 corresponding to which (3) had two linearly independent solutions of class (L^2). Correspondingly, every boundary condition (5) determines for (1) a spectrum $S(\phi)$, containing a (possibly vacuous) point spectrum $P(\phi)$. The latter consists of those values λ_0 corresponding to which there exists an $x = y(t)$ satisfying (3), (4), (5). A theorem of Weyl implies that the cluster values of $S(\phi)$, that is, of the set consisting of the continuous spectrum and the derivative of the point spectrum, is independent of ϕ; cf. [3], pp. 251-252. This invariant λ-set, $S'(\phi)$, can therefore be denoted

---

* Received July 2, 1948.

simply by S'. Accordingly, a λ is an isolated point of $S(\phi)$ if and only if it is in $P(\phi)$ without being in S'.

By an application of the method used in [7], pp. 27-30, it was shown in [2] that every real number not contained in S' is in $P(\phi)$ for some ϕ. By a further development of the same method, a method related to the Lebesgue-Toeplitz norm principle of *arbitrary* functions, a property of the eigenfunctions belonging to isolated points of the spectrum will be established in the sequel.

2. The definition, (4), of an eigenfunction does not imply that $y(t) \to 0$ as $t \to \infty$, and still less that $y(t)$ tends to 0 *fast*. Actually, even if (2) is refined to

$$f(t) \to 0 \text{ as } t \to \infty, \tag{6}$$

it is quite possible that (3) has for some λ_0, say for

$$\lambda_0 = 1, \tag{6 bis}$$

a solution $x = y(t)$ which, though it satisfies (4), fails to be $O(t^{-1})$ as $t \to \infty$; cf. [6], p. 269. On the other hand, the eigenfunctions tend very fast to 0 in all the "classical" cases of wave equations, that is to say, of differential equations (1) which are integrable "explicitly."

The purpose of the following consideration is to give some account of this situation. The explanation will be that in the "constructed" examples, in which an eigenfunction fails to tend to 0 *fast*, the corresponding eigenvalue cannot be an isolated point of the spectrum. What will actually be proved is the following theorem:

If a solution $x = y(t)$ *of* (3), *where* (2) *is assumed, satisfies* (4), *then either* $\lambda = \lambda_0$ *is not an isolated point of the spectrum or else*

$$y(t) = O(t^{-N}) \text{ for every } N \qquad (t \to \infty). \tag{7}$$

It remains undecided whether (7) can be refined to

$$y(t) = O(e^{-ct}) \text{ for some } c > 0. \tag{8}$$

In the explicit cases referred to above, as well as when $f(t)$ is periodic, (8) holds whenever (4) does.

If (7) could be refined to (8), at least when (2) is replaced by (6), there would result, as a corollary and with a direct proof, a theorem of Hartman [1], which states that (6 bis) cannot be an isolated point of the spectrum in the case (6). In fact, in order to obtain this as a corollary, it would be sufficient to observe that, as easily verified (cf., e. g., [4], p. 391), a solution of (3) in the case (6), (6 bis) must vanish identically if it is $O(e^{-\epsilon t})$ for some $\epsilon > 0$.

3. Let $x = y(t)$ be a function satisfying (4) and (3), where λ_0 is some real number and $f(t)$ is subject to (2). Suppose further that λ_0 is not in S', the common derivative of every $S(\phi)$. It will be proved that these assumptions imply (7). This will prove the italicized alternative (the claim of which is not disjunctive).

First, since $y(t)$ does not vanish identically, $y(0)$ and $y'(0)$ do not vanish simultaneously, and so (5) holds for $x = y$ and for a certain $\phi = \phi_0$ which is unique (mod π). Choose any ϕ satisfying

$$\phi \not\equiv \phi_0 \pmod{\pi}. \tag{9}$$

It is easy to see that λ_0 cannot then be in $S(\phi)$.

In fact, $S(\phi) = S' + P(\phi)$. But λ_0 is not in S', by assumption. Hence, λ_0 is not in $S(\phi)$ if it is not in $P(\phi)$. Suppose, if possible, that λ_0 is in $P(\phi)$. Then, since ϕ_0 was so defined as to make λ_0 a point of $P(\phi_0)$, it follows from (9) that two point spectra belonging to distinct boundary conditions have a point ($= \lambda_0$) in common. But this is impossible, since it means that (3) has two linearly independent solutions of class (L^2), which, as mentioned after (5), is prevented by (2).

This proves that λ_0 is not in $S(\phi)$. But to say that λ_0 is not in $S(\phi)$ is equivalent to the following assertion (cf. [3], p. 251): The inhomogeneous differential equation

$$v'' + (f(t) + \lambda_0)v = u(t) \tag{10}$$

and the boundary condition (5), where $x = v$, have a unique solution $v = v(t)$ satisfying

$$\int_0^\infty v^2(t)\,dt < \infty, \tag{11}$$

whenever $u(t)$ in (10), where $0 \leqq t < \infty$, is given so as to satisfy the following restrictions:

$$u(t) \text{ is continuous and } \int_0^\infty u^2(t)\,dt < \infty \tag{12}$$

(all functions are meant to be real-valued).

In particular, (10) has, for every continuous $u(t)$ of class (L^2), a solution $v(t)$ of class (L^2). Only this fact, which does not involve any boundary condition (and, correspondingly, no uniqueness) for $v(t)$, will be needed in the sequel.

4. Let $x = z(t)$ be any solution of (3) which is not a constant multiple of the solution $x = y(t)$ of (3). Then, since $y(t)$ does not vanish identically, the Wronskian of $y(t)$ and $z(t)$ is a non-vanishing constant. The latter can be made 1 if $z(t)$ is multiplied by a constant. Then

$$z(t)y'(t) - y(t)z'(t) = 1 \tag{13}$$

is an identity in t. Since both $x = y(t)$ and $x = z(t)$ satisfy (3), it is readily verified from (13) that, if $u(t)$ is any continuous function,

$$v_0(t) = y(t)\int_0^t z(s)u(s)\,ds - z(t)\int_0^t y(s)u(s)\,ds \tag{14}$$

is a solution, $v = v_0(t)$, of (10). Hence, every solution of (10) is of the form

$$v(t) = ay(t) + bz(t) + v_0(t), \tag{15}$$

where a, b are constants.

In view of (4) and of the linearity of the (L^2)-space, (15) will satisfy (11) for some a and some b only if (15) satisfies (11) for $a = 0$ and for some b. Hence, if (14) is substituted into (15), it is seen that (10) cannot have a solution $v(t)$ of class (L^2) for a given continuous $u(t)$ unless $v(t)$ can be chosen as

$$v(t) = bz(t) + y(t)\int_0^t z(s)u(s)\,ds - z(t)\int_0^t y(s)u(s)ds. \tag{16}$$

Since (10) is supposed to have a solution of class (L^2) whenever the given $u(t)$ satisfies (12), it follows that there belongs to every $u = u(t)$, satisfying (12), some constant $b = b_u$ corresponding to which the function (16) becomes of class (L^2). It will be assumed that the constant $b = b_u$ (which turns out to be unique) is so chosen.

In order to calculate $b = b_u$, differentiate (16), multiply the derived relation by $y(t)$, the relation (16) itself by $-y'(t)$, and add. This gives

$$yv' - vy' = b(z'y - z'y) + 0 - (z'y - y'z)\int_0^t u(s)y(s)\,ds,$$

an identity in t. It can be simplified to

$$w(t) = -b + \int_0^t u(s)y(s)\,ds, \tag{17}$$

if use is made of the identity (13) and of the abbreviation

$$w(t) = y(t)v'(t) - v(t)y'(t). \tag{18}$$

By Schwarz's inequality, (4) and (12) imply that the product uy is of class $(L) = (L^1)$. It follows therefore from (17) that $w(t)$ tends, as $t \to \infty$, to a limit, and that $w(\infty)$ is connected with the constant $b = b_u$ by the relation

$$b = -w(\infty) + \int_0^\infty u(s)y(s)\,ds, \tag{19}$$

where the integral is absolutely convergent.

5. It was shown in [5] that, if f, u satisfy (2), (12) respectively, then a solution, $v(t)$, of (10) cannot be of class (L^2) unless its derivative, $v'(t)$, is of class (L^2). Hence, not only (16) but also the derivative of (16) must be of class (L^2), if $u(t)$ is any function satisfying (12) and b in (16) denotes the constant (19). It also follows that $y'(t)$ is of class (L^2). In fact, $y(t)$ is of class (L^2) and is a solution of (3), i. e., of the case $u = 0$ of (12).

Accordingly, all four functions v, v', y, y' are of class (L^2). But (18) shows that w is a bilinear form in these functions. Consequently,

$$w(t) \text{ is of class } (L) = (L^1). \tag{20}$$

Since the limit $w(\infty)$ exists, it is clear from (20) that

$$w(\infty) = 0. \tag{21}$$

But (21) and (19) reduce (17) to

$$w(t) = -\int_t^\infty y(s)u(s)\,ds. \tag{22}$$

It follows therefore from (20) that

$$\int_0^\infty \left| \int_t^\infty y(s)u(s)\,ds \right| dt < \infty. \tag{23}$$

Accordingly, the eigenfunction $y(t)$ has the property that (23) holds for every $u(t)$ satisfying (12).

Since (12) is satisfied by $u=y$, it follows that

$$\int_0^\infty \left(\int_t^\infty y^2(s)\,ds \right) dt < \infty. \tag{24}$$

Hence, by Fubini's theorem,

$$\int_0^\infty y^2(s) \left(\int_0^s dt \right) ds < \infty; \tag{25}$$

so that

$$\int_0^\infty t y^2(t)\,dt < \infty. \tag{26}$$

6. In the above proof of (26), use had to be made of the fact that, due to (2),

$$\int_0^\infty y'^2(t)\,dt < \infty \tag{27}$$

is implied by (4). By adaptation of the proof of this implication (cf. [5]), it will now be shown that, due to (2),

$$\int_0^\infty t y'^2(t)\,dt < \infty \tag{28}$$

is implied by (26).

In terms of the given f of (1), define f^+ and f^- by placing

$$f^+(t) = \max\,(f(t), 0) \text{ and } f^-(t) = \min\,(f(t), 0).$$

Then $f^+ \geqq 0$, $f^- \leqq 0$, and $f = f^+ + f^-$. Since (2) means that $f^+(t) = O(1)$ as $t \to \infty$, it follows from (26) that

$$\int_0^t f^+(s) s y^2(s)\,ds = O(1).$$

Hence, if $x = y(t)$ and $f(t) = f^+(t) + f^-(t)$ are substituted into (3) and if the result is multiplied by $ty(t)$, it follows that

$$\int_0^t s y(s) y''(s)\,ds + \int_0^t s f^-(s) y^2(s)\,ds = O(1),$$

by (26). But

$$sy(s)y''(s) = \tfrac{1}{2}sy^{2''}(s) - sy'^2(s)$$

is an identity. Furthermore, a partial integration gives

$$\int_0^t \tfrac{1}{2}sy^{2''}(s)\,ds = ty(t)y'(t) - \int_0^t y(s)y'(s)\,ds.$$

Finally, from (4) and (27),

$$\int_0^t y(s)y'(s)\,ds = O(1),$$

by Schwarz's inequality. Clearly, the last four formula lines imply that

$$(29) \qquad ty(t)y'(t) - \int_0^t sy'^2(s)\,ds + \int_0^t sf^-(s)y^2(s)\,ds = O(1).$$

In order to deduce (28) from (29), suppose that (28) is false. Then, as $t \to \infty$,

$$-\int_0^t sy'^2(s)\,ds \to -\infty.$$

It follows therefore from (29), where $f^-(s) \leqq 0$, that $ty(t)y'(t) \to \infty$. In particular, $y(t)y'(t)$ is positive, hence $y^2(t)$ is increasing, from a certain t onward. Since this contradicts (4), the proof of (28) is complete.

A by-product of this consideration is that the eigenfunction $y(t)$ must have, with reference to the coefficient function of (1), the following property:

$$(30) \qquad \int_0^\infty t\,|\,f(t)\,|\,y^2(t)\,dt < \infty.$$

In fact, since (2) means that $f(t)$ is of the form $f^-(t) + O(1)$, where $f^-(t) \leqq 0$, it is clear from (26) that the negation of the truth of (30) is equivalent to

$$\int_0^t sf^-(s)y^2(s)\,ds \to -\infty.$$

On the other hand, from (29) and (28),

$$ty(t)y'(t) + \int_0^t sf^-(s)y^2(s)\,ds = O(1).$$

But the last two formula lines imply the relation $ty(t)y'(t) \rightarrow \infty$ which, as before, contradicts (4).

7. The assertion (20) depended on the fact that both $v(t)$ and $v'(t)$ are of class (L^2). On the other hand, (26) and (28) mean that both $t^{\frac{1}{2}}y(t)$ and $t^{\frac{1}{2}}y'(t)$ are of class (L^2). Hence (18) shows that $tw(t)$ is a bilinear form in functions of class (L^2), and so

$$tw(t) \text{ is of class } (L) = (L^1). \tag{31}$$

The proof of (26) depended entirely on (20). Correspondingly, if (20) is replaced by (31) in the deductions leading from (20) to (26), what results is the following refinement of (26):

$$\int_0^\infty t^2 y^2(t)\,dt < \infty.$$

This in turn leads to

$$\int_0^\infty t^2 y'^2(t)\,dt < \infty,$$

if the argument which led from (26) to (28) is repeated. In fact, (28) can now be used instead of (27).

Clearly, the process adapts itself to a complete induction and proves therefore that both

$$\int_0^\infty t^k y^2(t)\,dt < \infty \tag{32}$$

and

$$\int_0^\infty t^k y'^2(t)\,dt < \infty \tag{33}$$

hold for arbitrarily large values of k. It also follows that (30) can be replaced by

$$\int_0^\infty t^k \,|\, f(t) \,|\, y^2(t)\,dt < \infty. \tag{34}$$

Let $0 < t < T < \infty$ and $k > 1$. Since, by Schwarz's inequality,

$$|\, y(T) - y(t) \,|^2 \leqq \int_t^T s^k y'^2(s)\,ds \int_t^T s^{-k} ds,$$

(33) supplies the existence of a constant C_k satisfying

$$| y(T) - y(t) |) \leqq C_k \, | T^{1-k} - t^{1-k} |^{\frac{1}{2}}.$$

Hence, (7) follows by letting $T \rightarrow \infty$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman, "On the spectra of slightly disturbed linear oscillators," *American Journal of Mathematics*, vol. 71 (1949), pp. 71-79.

[2] ——— and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[3] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[4] A. Wintner, "The adiabatic linear oscillator," *American Journal of Mathematics*, vol. 68 (1946), pp. 385-397.

[5] ———, "(L^2)-connections between the potential and kinetic energies of linear systems," *ibid.*, vol. 69 (1947), pp. 5-13.

[6] ———, "Asymptotic integrations of the adiabatic oscillator," *ibid.*, vol. 69 (1947), pp. 251-272.

[7] ———, "On the location of continuous spectra," *ibid.*, vol. 70 (1948), pp. 22-30.

THE CLUSTER SPECTRA OF BOUNDED POTENTIALS.*

By C. R. Putnam.

---

1. Let $q = q(x)$ be a real continuous function on the half-line $0 \leqq x < \infty$ and let λ denote a real parameter. The differential equation

$$y'' + (\lambda - q)y = 0 \tag{1}$$

is said to be in the *Grenzpunktfall,* in the terminology of Weyl [9], p. 238, if for some λ (hence, for all λ) (1) possesses a solution y satisfying

$$\int_0^\infty y^2 dx = \infty.$$

Here, and in the sequel, only real functions will be considered.

In the *Grenzpunktfall,* the differential equation (1) and a homogeneous boundary condition

$$y(0) \cos \alpha + y'(0) \sin \alpha = 0, \qquad 0 \leqq \alpha < \pi, \tag{2}$$

determine a boundary value problem for every fixed α. It is known [9], p. 251, that the set, S', of the cluster points of the spectrum of any such boundary value problem is independent of α. Denote the least point (possibly $\pm \infty$) of S' by λ_0. It is known [5], p. 310, that the set of points constituting the spectrum of any fixed boundary value problem, determined by (1) and (2), cannot be bounded from above. Also, the set of points of the continuous spectrum either is empty or is unbounded; [12], p. 782.

If q is bounded,

$$| q | < \text{const.}, \qquad 0 \leqq x < \infty, \tag{3}$$

and even if q is just bounded from below,

$$q > \text{const.}, \qquad 0 \leqq x < \infty, \tag{3 bis}$$

the differential equation (1) is in the *Grenzpunktfall*; [9], p. 238. According to the corollary in [3], p. 850, the condition (3) implies that λ_0 satisfies the inequality $A \leqq \lambda_0 \leqq B$, where A and B are defined by

---

* Received December 11, 1948.

$$(4) \qquad A = \liminf_{x\to\infty} q(x), \qquad B = \limsup_{x\to\infty} q(x).$$

The theorem to be proved in the present paper yields an extension of these results to information concerning the set S'. It will be shown that *if q satisfies* (3), then a "gap" in S' (that is, a λ-interval occurring in the decomposition of the (possibly empty) open set constituting the complement of S' with respect to the half-line $\lambda_0 \leqq \lambda < \infty$) cannot have a length greater than $B - A$. In particular, condition (3) implies that the set S' is unbounded. The milder restriction (3 bis) surely is not sufficient to guarantee this result; in fact, according to [9], p. 252, if $q \to \infty$ as $x \to \infty$, the set S' is empty.

All of this is contained in the following theorem:

(*) *If $q = q(x)$ is continuous on $0 \leqq x < \infty$ and satisfies* (3), *then every λ-interval, $\mu_1 < \lambda < \mu_2$, contains at least one point of S' whenever the length of the interval satisfies*

$$(5) \qquad \mu_2 - \mu_1 > \limsup_{x\to\infty} q(x) - \liminf_{x\to\infty} q(x)$$

and the lower end point μ_1 satisfies

$$\mu_1 \geqq \liminf_{x\to\infty} q(x).$$

As a consequence of (*), one obtains the theorem [2], p. 71, that if $q(x) \to 0$ as $x \to \infty$, then every point λ of the half-line $0 \leqq \lambda < \infty$ belongs to S'.

It is known [9], p. 251, that a point λ is not in the spectrum of the boundary value problem determined by (1) and (2), for a fixed α, if and only if for *every* continuous function $f(x)$ of class (L^2) on $0 \leqq x < \infty$ the inhomogeneous differential equation

$$(6) \qquad y'' + (\lambda - q)y = f$$

possesses a unique solution of class (L^2) satisfying the boundary condition (2). The proof of (*) will depend upon this characterization of the spectrum together with an adaptation of the principle (which can be regarded as a manifestation of the Lebesgue-Toeplitz norm construction) used in [11], pp. 26-27, and then applied in [4], [8] and [2], for the location of points of the spectrum.

2. *Proof of* (*). Let $\phi = \phi(x, \lambda, \alpha)$ denote that solution of (1) satisfying (2) for which

$$\phi(0, \lambda, \alpha) = -\sin \alpha \text{ and } \phi'(0, \lambda, \alpha) = \cos \alpha,$$

where the prime denotes partial differentiation with respect to x. Let $\lambda_1, \lambda_2, \cdots$ denote the (possibly empty) set of points of the point spectrum of the boundary value problem determined by (1') and (2). If ϕ_j denotes the normalized eigenfunction belonging to the eigenvalue λ_j, then

$$L(\phi_j) + \lambda_j \phi_j = 0, \tag{7}$$

where $L(y)$ denotes the differential operator

$$L(y) = y'' - qy. \tag{8}$$

According to Weyl's application [9], pp. 239-251, of Hellinger's spectral resolution theory [6] to differential equations, there exists a continuous monotone function $\rho = \rho(\mu)$ such that the function $P = P(x)$,

$$P(x) = P(x, \lambda, \alpha) = \int_0^\lambda \phi(x, \mu, \alpha)\, d\rho(\mu),$$

is of class (L^2) and satisfies

$$L(P(x, \lambda, \alpha)) + \lambda P(x, \lambda, \alpha) = \int_0^\lambda P(x, \mu, \alpha)\, d\mu. \tag{9}$$

In terms of the eigendifferentials $d_\lambda P = dP(x, \lambda, \alpha)$ the differential equation (9) may be expressed as

$$L(dP) + \lambda dP = 0. \tag{9 bis}$$

The eigenfunctions ϕ_j together with the eigendifferentials dP constitute an orthonormal complete set of functions on $0 \leqq x < \infty$; hence, if y is any function of class (L^2) on $0 \leqq x < \infty$, with "Fourier coefficients"

$$c_j = \int_0^\infty y(x) \phi_j(x)\, dx, \qquad \Delta\Gamma(\lambda) = \int_0^\infty y(x) \Delta P(x, \lambda, \alpha)\, dx, \tag{10}$$

the Parseval relation

$$\int_0^\infty y^2(x)\, dx = \int_{-\infty}^\infty (d\Gamma)^2/d\rho + \sum_j c_j^2 \tag{11}$$

holds.

Since the addition of a constant c to $q(x)$ merely translates the spectrum, and hence S', by c, it can be supposed, without loss of generality, that

$$0 \leqq B = -A < \infty \tag{12}$$

(cf. (4)). The proof of (*) will be complete if it is shown that the λ-interval $\mu_1 < \lambda < \mu_2$, where $\mu_1 \geqq -B$, contains at least one point of S' whenever $\mu_2 - \mu_1 > 2B$.

Suppose, if possible, that the interval $\mu_1 < \lambda < \mu_2$ contains no points of S'. Choose an arbitrary positive number ϵ so small that $\mu_1 + 2\epsilon < \mu_2 - 2\epsilon$ and that

$$\lambda^* = (\mu_1 + \mu_2)/2 + \epsilon \tag{13}$$

is not an eigenvalue. It is clear that $\lambda^* > 0$. Denote by $\lambda^1, \lambda^2, \cdots, \lambda^M$, where $M = M_\epsilon$, the eigenvalues (if any) satisfying

$$\mu_1 + \epsilon \leqq \lambda^j \leqq \mu_2 - \epsilon, \qquad j = 1, 2, \cdots, M, \tag{14}$$

and denote by $\phi^1, \phi^2, \cdots, \phi^M$ the corresponding normalized eigenfunctions.

In the sequel, the symbol f will always denote a continuous function, of class (L^2) on $0 \leqq x < \infty$, satisfying

$$\int_0^\infty f^2 dx = 1 \tag{15}$$

and

$$\sum_{j=1}^{M} \left(\int_0^\infty f\phi^j dx / (\lambda^* - \lambda^j) \right)^2 < \epsilon. \tag{16}$$

Any statement involving "independent of f" will mean independent of the choice of f in the class of continuous functions f for which (15) and (16) hold. Consider the inhomogeneous differential equation (6) for $\lambda = \lambda^*$, that is, by (8),

$$L(y) + \lambda^* y = f. \tag{17}$$

Since λ^* does not belong to the spectrum of the boundary value problem determined by (1) and (2), there exists a unique solution y of class (L^2) satisfying (2), cf. [9], p. 251. It follows from the Green identity (cf. [9], pp. 241-242) that

$$\int_0^\infty yL(\phi_j)\,dx = \int_0^\infty L(y)\phi_j dx \text{ and } \int_0^\infty yL(\Delta P)\,dx = \int_0^\infty L(y)\,\Delta P dx.$$

Moreover, one obtains from (7) and the first relation of (10)

$$\int_0^\infty yL(\phi_j)\,dx = -\lambda_j c_j$$

and from (9) or (9 bis) and the second relation of (10)

$$\int_0^\infty yL(\Delta P)\,dx = -\int_\lambda^{\lambda+\Delta\lambda} \mu\, d\Gamma(\mu).$$

(Cf. [7], pp. 782-785; there, however, the contributions of both the point spectrum and the continuous spectrum are included in the function $P(x, \lambda, \alpha)$.) It follows from (17) and the last four formula lines that

$$(18) \qquad \int_0^\infty f\phi_j\,dx = (\lambda^* - \lambda_j)\,c_j$$

and

$$(19) \qquad \int_0^\infty f\Delta P\,dx = \int_\lambda^{\lambda+\Delta\lambda} (\lambda^* - \mu)\,d\Gamma(\mu).$$

If η is defined by

$$(20) \qquad \eta = (\mu_2 - \mu_1)/2 - 2\epsilon,$$

then relations (18), (19) and the Parseval relation for f yield the inequality

$$(21) \qquad \int_0^\infty f^2\,dx \geqq \eta^2\Big(\int^* (d\Gamma)^2/d\rho + \Sigma^* c_j{}^2\Big),$$

where $\Delta\Gamma$ and c_j are defined by (10) and where the asterisk indicates that the integration and summation are to be taken over that part of the spectrum exterior to the λ-interval $\mu_1 + \epsilon \leqq \lambda \leqq \mu_2 - \epsilon$. Since the only contribution of this interval to the spectrum is that due to the points $\lambda^1, \lambda^2, \cdots, \lambda^M$ (cf. (14)), relations (11), (16), (18) and (21) imply that

$$\int_0^\infty f^2\,dx \geqq \eta^2\Big(\int_0^\infty y^2\,dx - \epsilon\Big).$$

Thus, in virtue of (15) and (20)

$$(22) \qquad \eta^2 \int_0^\infty y^2\,dx \leqq 1 + \epsilon\eta^2 < C_\epsilon$$

where the constant C_ϵ depends on ϵ but is independent of f (and y). The inequality (22) will play a fundamental role in the calculations to follow.

Henceforth, it will be supposed that the boundary value problem under consideration is that determined by the differential equation (1) and the boundary condition (2) for $\alpha = 0$, that is, $y(0) = 0$. This is no loss of generality since S' is independent of the choice of α in (2). Let the function $\psi = \psi(x)$ on $0 \leqq x < \infty$ be defined by

$$(23) \qquad \psi = \sin(\lambda^{*\frac{1}{2}} x),$$

so that ψ is real since $\lambda^* > 0$. The next step in the proof of (*) will be to show that there exists a number $N = N_\epsilon$, depending only on ϵ and λ^* and *not* on f or y, having the property that on any x-interval of length N there exists at least one point ξ such that

$$(24) \qquad | y'(\xi)\psi(\xi) - y(\xi)\psi'(\xi) | < \epsilon.$$

In virtue of (17),

$$-\int_0^X yL(y)\,dx = \lambda^* \int_0^X y^2 dx - \int_0^X fy dx,$$

while an integration by parts yields

$$-\int_0^X yL(y)\,dx = -yy' \Big|_0^X + \int_0^X (y'^2 + qy^2)\,dx.$$

Since the functions y and f belong to class (L^2) and q satisfies condition (3), it follows from the proof of the theorem [10], p. 6 (cf. also [1]), that $y(X)y'(X) \to 0$ as $X \to \infty$. Since $y(0) = 0$, one obtains from the last two formula lines that

$$(25) \qquad \int_0^\infty (y'^2 + qy^2)\,dx = \lambda^* \int_0^\infty y^2 dx - \int_0^\infty fy dx.$$

An application of the Schwarz inequality to the second integral on the right of the last inequality yields, in virtue of (15),

$$\left| \int_0^\infty fy dx \right| \leqq \left(\int_0^\infty y^2 dx \right)^{\frac{1}{2}}.$$

Relations (25), (3) and (22) imply

$$(26) \qquad \int_0^\infty y'^2 dx < D_\epsilon$$

where the constant D_ϵ is again independent of f. The existence of the number N is clearly a consequence of (22), (23) and (26).

It will be shown that in virtue of the above results it is possible to define a continuous function $f(x)$ on $0 \leqq x < \infty$, satisfying (15) and (16), and to select a pair of positive constants $\xi_1 < \xi_2$ in such a way that the function y (which depends on the choice of f) satisfies the inequality

$$(27) \qquad (B + \epsilon) \left(\int_{\xi_1}^{\xi_2} y^2 dx \right)^{\frac{1}{2}} \geqq (1 - \epsilon)^{\frac{1}{2}} - 2\epsilon.$$

Suppose this has been proved. Then, by the first inequality of (22),

$$(B+\epsilon)(1+\epsilon\eta^2)^{\frac{1}{2}}/\eta \geqq (1-\epsilon)^{\frac{1}{2}} - 2\epsilon.$$

Since ϵ can be chosen arbitrarily small, the last inequality and (20) imply $2B/(\mu_2-\mu_1) \geqq 1$, that is,

$$\mu_2 - \mu_1 \leqq 2B. \tag{28}$$

Since (28) is in contradiction with the assumption $\mu_2-\mu_1 > 2B$, the supposition that the λ-interval $\mu_1 < \lambda < \mu_2$ contains no points of S' is untenable.

Thus, in order to complete the proof of (*), it remains only to establish relation (27) for suitably chosen f and ξ_1, ξ_2. In virtue of (12) and (4) there exists a positive number $T = T_\epsilon$ so large that

$$\left(\int_a^b q^2\psi^2 dx \Big/ \int_a^b \psi^2 dx\right)^{\frac{1}{2}} \leqq B+\epsilon \text{ whenever } b > a \geqq T. \tag{29}$$

In addition, T can be chosen so that

$$\psi(T) = 0 \tag{30}$$

and

$$\sum_{j=1}^{M} \int_T^\infty (\phi^j)^2 dx/(\lambda-\lambda^j)^2 < \epsilon. \tag{31}$$

On the other hand, (23) shows that $\int_0^\infty \psi^2 dx = \infty$. Hence, if T is fixed,

$$\int_{T+N}^{U-N} \psi^2 dx \Big/ \int_T^U \psi^2 dx > 1-\epsilon \tag{32}$$

whenever $U = U_\epsilon$ ($> T+2N$) is sufficiently large. Let U ($> T+2N$) be chosen so as to satisfy (32), $\psi(U) = 0$ and

$$\int_{T+N}^{U-N} \psi^2 dx \geqq 1. \tag{33}$$

Define the function $f(x)$ on $0 \leqq x < \infty$ by

$$f = \begin{cases} \psi/\left(\int_T^U \psi^2 dx\right)^{\frac{1}{2}} & \text{if } T \leqq x \leqq U, \\ 0 & \text{if } 0 \leqq x \leqq T \text{ and } U \leqq x < \infty. \end{cases}$$

It follows from (30) and the definition of U that f is continuous on $0 \leqq x < \infty$ and satisfies (15). That f also satisfies (16) is a consequence of (31) and (15). Finally, the property of the number N shows that there exist numbers ξ_1 and ξ_2 satisfying

$$T \leqq \xi_1 \leqq T + N, \; U - N \leqq \xi_2 \leqq U \tag{34}$$

and (24) for $\xi = \xi_1$ and $\xi = \xi_2$. It is clear from (32), the definition of U and (34) that

$$\int_{\xi_1}^{\xi_2} \psi^2 dx / \int_T^U \psi^2 dx > 1 - \epsilon. \tag{35}$$

Also, (33) and (34) imply

$$\int_{\xi_1}^{\xi_2} \psi^2 dx \geqq 1. \tag{36}$$

Since ψ satisfies the differential equation

$$\psi'' + \lambda^* \psi = 0,$$

it follows from (17) that

$$y''\psi - y\psi'' = qy\psi + f\psi$$

and, consequently,

$$[y'\psi - y\psi']_{\xi_1}^{\xi_2} = \int_{\xi_1}^{\xi_2} qy\psi dx + \int_{\xi_1}^{\xi_2} f\psi dx.$$

Hence, in virtue of (24) for $\xi = \xi_1$ and $\xi = \xi_2$,

$$\left| \int_{\xi_1}^{\xi_2} qy\psi dx \right| \geqq \int_{\xi_1}^{\xi_2} f\psi dx - 2\epsilon;$$

but, from (35) and the definition of f, it follows that

$$\int_{\xi_1}^{\xi_2} f\psi dx \geqq (1 - \epsilon)^{\frac{1}{2}} \left(\int_{\xi_1}^{\xi_2} \psi^2 dx \right)^{\frac{1}{2}}.$$

The last two relations and (36) show that

$$\left| \int_{\xi_1}^{\xi_2} qy\psi dx \right| \geqq \left((1 - \epsilon)^{\frac{1}{2}} - 2\epsilon \right) \left(\int_{\xi_1}^{\xi_2} \psi^2 dx \right)^{\frac{1}{2}}.$$

Finally, the Schwarz inequality and relations (29) and (34) give

$$\left| \int_{\xi_1}^{\xi_2} qy\psi dx \right| \leqq (B + \epsilon) \left(\int_{\xi_1}^{\xi_2} \psi^2 dx \right)^{\frac{1}{2}} \left(\int_{\xi_1}^{\xi_2} y^2 dx \right)^{\frac{1}{2}}.$$

The last two formula lines imply (27), and so the proof of (*) is complete.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman, "The L^2-solutions of linear differential equations of second order," *Duke Mathematical Journal*, vol. 14 (1947), pp. 323-326.

[2] ———, "On the spectra of slightly disturbed linear oscillators," *American Journal of Mathematics*, vol. 71 (1949), pp. 71-79.

[3] ——— and C. R. Putnam, "The least cluster point of the spectrum of boundary problems," *ibid.*, vol. 70 (1948), pp. 849-855.

[4] ——— and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[5] ——— and A. Wintner, "On the orientation of unilateral spectra," *American Journal of Mathematics*, vol. 70 (1948), pp. 309-316.

[6] E. Hellinger, "Neue Begründung der Theorie quadratischer Formen von unendlicher Veränderlichen," *Journal für die reine und angewandte Mathematik*, vol. 136 (1909), pp. 210-271.

[7] C. R. Putnam, "An application of spectral theory to a singular calculus of variations problem," *American Journal of Mathematics*, vol. 70 (1948), pp. 780-803.

[8] ———, "On the spectra of certain boundary value problems." *ibid.*, vol. 71 (1949), pp. 109-111.

[9] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 220-269.

[10] A. Wintner, "(L^2)-connections between the potential and kinetic energies of linear systems," *American Journal of Mathematics*, vol. 69 (1947), pp. 5-13.

[11] ———, "On the location of continuous spectra," *ibid.*, vol. 70 (1948), pp. 22-30.

[12] ———, "On Dirac's theory of continuous spectra," *Physical Review*, vol. 73 (1948), pp. 781-785.

ON THE NON-VANISHING AT $s=1$ OF CERTAIN DIRICHLET SERIES.*

By GEORGE SHAPIRO.

By a multiplicative function is meant a function, $f(n)$, which is defined on the set of positive integers, is not identically zero, and satisfies $f(nm)=f(n)f(m)$ if $(n,m)=1$. If the last proviso is unnecessary, that is, if $f(nm)=f(n)f(m)$ for all n and m, then $f(n)$ is called a completely multiplicative function.

The restriction that $f(n)$ is not identically zero is equivalent to $f(1)=1$. Therefore any completely multiplicative function may be obtained by choosing $f(p)$ for each prime, p, and by setting

$$(1)\qquad f(n)=f^{k_1}(p_1)\cdots f^{k_m}(p_m) \text{ if } n=p_1^{k_1}\cdots p_m^{k_m} \text{ and } f(1)=1.$$

Any multiplicative function may be obtained by choosing $f(p^k)$ for each prime, p, and each positive integer, k, and setting.

$$(2)\qquad f(n)=f(p_1^{k_1})\cdots f(p_m^{k_m}) \text{ if } n=p_1^{k_1}\cdots p_m^{k_m} \text{ and } f(1)=1.$$

Formally, (1) and (2) are equivalent to the Euler relations

$$(3)\qquad \sum f(n)/n^s=\prod_p (1-f(p)/p^s)^{-1}$$

and

$$(4)\qquad \sum f(n)/n^s=\prod_p \left(\sum_{\nu=0}^{\infty} f(p^\nu)/p^{\nu s}\right),$$

respectively, where $\sum=\sum_{n=1}^{\infty}$ in (3) and (4), and in the sequel. The statement that $f(n)$ is a bounded completely multiplicative function is equivalent to $|f(n)|\leqq 1$. In this case, (3) is valid for $\Re s>1$ by virtue of absolute convergence.

If $f(n)$ is a residue character mod m, then the non-vanishing of $L(s)=\sum f(n)/n^s$ at $s=1$ is of prime importance in the proof of Dirichlet's theorem that infinitely many primes occur in the sequence $h, h+m, h+2m, \cdots$, where $(h,m)=1$. In this case, however, the absolute value of $f(n)$ is either 0 or 1.

---

* Received December 1, 1948.

Ingham [1] gave a simple direct proof that $L(1) \neq 0$ if $f(n)$ is any residue character; previously the case in which $f(n)$ assumes only the values $0, +1, -1$, required separate consideration. Wintner [3] extended Ingham's method to bounded, real, completely multiplicative functions and, in fact, to those satisfying $-1 \leqq f(p)$ for all p. In this paper, a combination of the methods of Ingham [1] and Wintner [3] is used to generalize Wintner's result to complex functions. In Theorem I below, Wintner's result for real bounded functions is extended to complex bounded functions. In Theorem II the boundedness condition is weakened in certain directions; however, Wintner's result is not obtained as a corollary.

THEOREM I. *If $f(n)$ is a bounded completely multiplicative function, then $L(s) = \sum f(n)/n^s$ is absolutely convergent for $\sigma = \Re(s) > 1$. If $L(s)$ possesses a regular analytic continuation onto a domain containing the real segment $\frac{1}{2} \leqq \sigma \leqq 1$, then $L(1) \neq 0$.*

Proof. Let

$$E(n) = \sum_{d \mid n} f(d).$$

Then $\sum | E(n) |^2/n^s$ is absolutely convergent for $\sigma > 1$ since $| E(n) | \leqq d(n)$, the number of divisors of n. Let $J(s) = \sum | f(n) |^2/n^s$, so that the corresponding Euler relation (3) implies

$$J(s) = \prod_p (1 - | f(p) |^2/p^s)^{-1} \text{ if } \sigma > 1.$$

It follows that if $\sigma > 1$, then

$$\sum | E(n) |^2/n^s = \bar{L}(s) L(s) \zeta(s) J(s)/J(2s); \tag{5}$$

cf., e. g. [1].

If $H(n)$ is defined by

$$H(n) = 1 \text{ or } H(n) = \prod_{p \mid n} (1 - | f(p) |^2)$$

according as $n = 1$ or $n > 1$, then for a fixed prime p,

$$(1 - | f(p) |^2/p^s)(1 - 1/p^s)^{-1} = 1 + \sum_{\nu=1}^{\infty} (1 - | f(p) |^2)/p^{\nu s} = \sum_{\nu=0}^{\infty} H(p^\nu)/p^{\nu s}.$$

Clearly, $H(n)$ is multiplicative and satisfies $0 \leqq H(n) \leqq 1$; hence, for $\sigma > 1$,

$$\sum H(n)/n^s = \prod_p (1 - | f(p) |^2/p^s)(1 - 1/p^s)^{-1} = \zeta(s)/J(s).$$

Multiplying the last relation by (5), we obtain

$$(6)\qquad \zeta^2(s)\bar{L}(s)L(s)/J(2s) = (\sum H(n)/n^s)(\sum |E(n)|^2/n^s) \text{ for } \sigma > 1.$$

The two series on the right are absolutely convergent and so their product may be represented by a Dirichlet series, $\sum G(n)/n^s$, which also converges absolutely for $\sigma > 1$. Clearly, $G(1) = 1$ and $G(n) \geqq 0$.

Now $J(s)$ is regular and does not vanish for $\sigma > 1$; therefore the same is true of $J(2s)$ when $\sigma > \frac{1}{2}$. The function $\zeta^2(s)$ is regular for $\sigma > \frac{1}{2}$ except for a double pole at $s = 1$. If we assume that $L(1) = 0$, then $\bar{L}(1) = 0$ and the pole of $\zeta^2(s)$ at $s = 1$ is absorbed in the product on the left of (6). Consequently, the function on the left of (6) is regular for $\sigma > \frac{1}{2}$, and by Landau's extension [2] of the Vivanti-Pringsheim theorem, the series $\sum G(n)/n^s$ converges absolutely and represents the function on the left of (6) for $\sigma > \frac{1}{2}$.

Now two cases arise,

$$\text{(i)}\qquad \sum |f(n)|^2/n < \infty$$

or

$$\text{(ii)}\qquad \sum |f(n)|^2/n = \infty.$$

If (i) is true, then $\sum |f(p)|^2/p < \infty$ and $\sum' 1/p < \infty$, where $'$ indicates summation over those primes for which $|f(p)| \geqq \frac{1}{2}$. Also (6) can be written as

$$(7)\qquad \zeta^2(s)\bar{L}(s)L(s) = J(2s) \sum G(n)/n^s,$$

where the two Dirichlet series on the right are absolutely convergent for $\sigma > \frac{1}{2}$, so that their product may be represented there by an absolutely convergent Dirichlet series, say $\sum e(n)/n^s$. Thus

$$\sum e(n)/n^s = (\sum 1/n^s)^2(\sum f(n)/n^s)(\sum \bar{f}(n)/n^s),$$

where

$$e(n) = \sum_{k|n} d(n/k)(\sum_{m|k} f(m)\bar{f}(k/m));$$

in particular, $e(p) = 2 + 2\Re f(p)$. Now $\sum |e(p)|/p < \infty$; and if $|f(p)| < \frac{1}{2}$, then $|e(p)| \geqq 1$, so that $\sum'' 1/p < \infty$, where the $''$ indicates summation over those primes for which $|f(p)| < \frac{1}{2}$. Hence, if (i) holds, we obtain the contradictory result $\sum 1/p < \infty$.

The case (ii) will now be considered. Since $G(1) = 1$ and $G(n) \geqq 0$, the relation (7) implies

$$|\zeta(\sigma)L(\sigma)|^2 \geqq |J(2\sigma)| \text{ for } \sigma > \tfrac{1}{2}.$$

But this again leads to a contradiction since $|\zeta(\sigma)L(\sigma)|^2$ has a finite limit as σ tends to $\frac{1}{2}+0$, whereas (ii) implies that $|J(2\sigma)| \to \infty$ as $\sigma \to \frac{1}{2}+0$.

Consequently, the assumption $L(1)=0$ is untenable and the proof of Theorem I is complete.

A somewhat different treatment is required in order to extend the result to unbounded completely multiplicative functions. We first prove two lemmas concerning the non-vanishing of the Dirichlet series with coefficients $\mu^2(n)f(n)$, where $\mu(n)$ is the Möbius function.

LEMMA A. *Let $g(n)$ be a multiplicative function with the properties that*

$$g(p^k)=0 \text{ if } k>1 \tag{8}$$

and

$$\Re g(p) \geqq -\tfrac{1}{2} \text{ for all primes } p. \tag{9}$$

If $G(s)=\sum g(n)/n^s$ converges absolutely on some half plane $\sigma>\sigma_0$ (>1) and possesses a regular analytic continuation onto a domain containing the real segment $1 \leqq \sigma \leqq \sigma_0$, then $G(1) \neq 0$.

Proof. First, $\bar{G}(s)G(s)\zeta(s)=\sum N(n)/n^s$ for $\sigma>\sigma_0$, where $N(n)$ is the multiplicative function,

$$N(n)=\sum_{k\mid n} g(n/k) \sum_{m\mid k} \bar{g}(m);$$

so that $N(p)=1+2\Re g(p)$ and

$$N(p^k)=(1+g(p))(1+\bar{g}(p)) \geqq 0 \text{ if } k>1.$$

Since (9) implies $N(p) \geqq 0$ as well, $N(n) \geqq 0$ for all n.

Now if $G(1)=0$, then $\bar{G}(\sigma)G(\sigma)\zeta(\sigma)$ is regular for $\sigma \geqq 1$ and vanishes at $\sigma=1$. By Landau's extension [2] of the Vivanti-Pringsheim theorem, the series $\sum N(n)/n^s$ converges and represents this function in this region. In particular, $\sum N(n)/n=0$, which is clearly impossible.

LEMMA B. *Let $g(n)$ be a multiplicative function with the properties that*

$$g(p^k)=0 \text{ if } k>1 \tag{8}$$

and

$$\Re g(p) \geqq -1 \text{ and } 2+4\Re g(p)+|g(p)|^2 \geqq 0 \text{ for all primes } p. \tag{10}$$

If $G(s)=\sum g(n)/n^s$ converges absolutely on some half plane $\sigma>\sigma_0$ $(>\frac{1}{2})$

and possesses a regular analytic continuation onto a domain containing the real segment $\frac{1}{2} \leqq \sigma \leqq \sigma_0$, *then* $G(1) \neq 0$.

Proof. First, $\zeta^2(s)/\zeta(2s) = \sum 2^{\nu(n)}/n^s$ for $\sigma > 1$, where $\nu(n)$ is the number of distinct prime divisors of n. Hence

$$\bar{G}(s)G(s)\zeta^2(s)/\zeta(2s) = \sum N(n)/n^s \text{ for } \sigma > \sigma_0, \tag{11}$$

where $N(n)$ is the multiplicative function determined by

$$N(p) = 2 + 2\Re g(p),\, N(p^2) = 2 + 4\Re g(p) + |\, g(p)\,|^2$$

and, if $k > 2$,

$$N(p^k) = 2 + 4\Re g(p) + 2\,|\, g(p)\,|^2 \geqq 0.$$

The condition (10) assures that $N(n) \geqq 0$ for all n.

If $G(1) = 0$, the function on the left of (11) is regular on the real half-line $\sigma > \frac{1}{2}$; by the theorem referred to before in [2], the series on the right converges and represents this function there. Then

$$|\, G(\sigma)\zeta(\sigma)\,|^2 = |\,\zeta(2\sigma)\,| \sum N(n)/n^\sigma \geqq |\,\zeta(2\sigma)\,|$$

for $\sigma > \frac{1}{2}$. But this is impossible, for the left side is bounded while the right side tends to infinity as σ tends to $\frac{1}{2} + 0$.

An application of either of these lemmas allows the boundedness assumption on $f(n)$ in Theorem I to be weakened in certain directions; and in the case of an application of Lemma A, a relaxation of the assumption concerning the analytic continuability of $L(s)$ is also possible.

THEOREM II. *Let* $f(n)$ *be a completely multiplicative function with the properties that* $\sum f(n)/n^s$ *converges absolutely on some half-plane* $\sigma > \sigma_0$ (> 1), *that* $\sum f(n^2)/n^2$ *converges and is not* 0, *and that* $g(n) = \mu^2(n)f(n)$ *satisfies either* (9) *or* (10) *(as well as* (8)*). Then* $G(s) = \sum g(n)/n^s$ *converges absolutely for* $\sigma > \sigma_0$. *Suppose, further, that* $G(s)$ *satisfies the assumption of analyticity made in either Lemma* A *or* B, *according as* $g(n)$ *satisfies* (9) *or* (10). *Then* $L(s)$ *possesses a regular analytic continuation onto a domain containing the real segment* $1 < \sigma \leqq \sigma_0$, *and* $L(1 + 0)$ *exists and is not zero.*

Proof. Let $r(n)$ be $f(n)$ or 0 according as n is or is not a square. Then $r(n)$ is multiplicative and the series $\sum r(n)/n^s$ converges absolutely for $\sigma > \sigma_0$. Furthermore,

$$(\sum g(n)/n^s)(\sum r(n)n/^s) = \sum f(n)/n^s,$$

since

$$g(1)r(p^k) + g(p)r(p^{k-1}) = f(p^k) \qquad (k = 1, 2, \cdots).$$

Thus $L(s) = G(s) \sum f(n^2)/n^{2s}$ for $\sigma > \sigma_0$; so that $L(s)$ can be continued analytically onto a domain containing the real segment $1 < \sigma \leqq \sigma_0$, while $L(1+0)$ exists and

$$L(1+0) = G(1) \sum f(n^2)/n^2.$$

Hence $L(1+0) \neq 0$, since $G(1) \neq 0$ by Lemma A or B, while the second factor in the last formula line does not vanish by assumption.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. E. Ingham, "Note on Riemann's zeta function and Dirichlet's L-functions," *Journal of the London Mathematical Society*, vol. 5 (1930), pp. 107-112.

[2] E. Landau, "Über einen Satz von Tschebyschef," *Mathematische Annalen*, vol. 61 (1905), pp. 527-550.

[3] A. Wintner, "The fundamental lemma in Dirichlet's theory of the arithmetical progressions," *American Journal of Mathematics*, vol. 68 (1946), pp. 385-392.

OSCILLATORY AND NON-OSCILLATORY LINEAR DIFFERENTIAL EQUATIONS.*

By Philip Hartman and Aurel Wintner.

In the differential equation

$$x'' + f(t)x = 0, \tag{1}$$

let $f(t)$ be real-valued and continuous for large positive t, say for

$$0 \leqq t < \infty, \tag{2}$$

and consider only those solutions $x(t)$ of (1) which are real-valued and do not vanish identically. These obvious restrictions of the coefficient function, $f(t)$, and of the admitted solutions, $x(t)$, of (1) will not be repeated in the wording or the proof of the results below.

The object will be to delimit, on the one hand, general possibilities for the behavior either of some or of all (i. e., either of one or of two linearly independent) solutions of (1) as $t \rightarrow \infty$ and to obtain, on the other hand, corresponding results for spectra (in Hilbert's sense). The latter aspect results if $f(t)$ in (1) is replaced by $f(t) + \lambda$, where λ is a real parameter, and an homogeneous linear boundary condition is assigned at $t = 0$.

As the various possibilities lead into quite different directions, each of the chapters will have its own introduction, explaining the problem at hand.

The Oscillatory Case.

If $f(t)$ is a positive constant, then every solution $x(t)$ of (1), being a sine curve, is $O(1)$ as $t \rightarrow \infty$. On the other hand, if $f(t) = \text{const.} > 0$ is relaxed to the assumption that

$$\lim_{t\rightarrow\infty} f(t) \text{ exists and is positive } (\neq \infty), \tag{3}$$

then, as shown by Perron [14], it is possible for (1) to have a solution for which

$$\limsup_{t\rightarrow\infty} | x(t) | = \infty. \tag{4}$$

---

* Received October 2, 1948.

From his counter-example, it is not possible to answer *either* of the following questions: If (3) is assumed, can (4) hold (α) for some but not for every, (β) for every, solution $x(t)$ of (1)? The first part of (i) below answers both of these questions in the affirmative. The second part of (i) represents an unexpected residue of the $O(1)$-situation prevailing in the trivial case $f(t) = \text{const.} > 0$.

(i) *Under the assumption* (3),

(α) *some but not every,* (β) *every,*

solution $x(t)$ *of* (1) *can fail to satisfy*

$$x(t) = O(1) \text{ as } t \rightarrow \infty. \tag{5}$$

On the other hand, (3), *and even the more general assumption*

$$0 < \liminf_{t \rightarrow \infty} f(t) \leqq \infty, \tag{6}$$

implies the existence of a sequence of t-intervals (t'_n, t''_n) *which tend to* $t = \infty$ *as* $n \rightarrow \infty$, *are independent of the choice of the pair of integration constants determining a solution of* (1), *and have the following property: If* t^* *varies on the (unbounded, open) t-set*

$$\sum_{n=1}^{\infty} (t'_n, t''_n), \tag{7}$$

a set depending only on f, then

$$x(t^*) = O(1) \text{ as } t^* \rightarrow \infty \tag{8}$$

holds for every solution $x(t)$ *of* (1).

If the assumption (3) is replaced by its limiting case,

$$f(t) \rightarrow \infty \text{ as } t \rightarrow \infty, \tag{9}$$

and *if the "growth" of* $f(t)$ *is smooth enough,* it follows from asymptotic formulae that

$$x(t) \rightarrow 0 \text{ as } t \rightarrow \infty \tag{10}$$

holds for every solution of (1); cf. [21]. But if the last italicized proviso is omitted, i. e., if only (9) is assumed, then not even (5) can in general be asserted. For an "explicit" example of such an $f(t)$, consisting of a transfer of Perron's construction from (3) to (9), cf. [22].

This leaves, however, undecided, for (9), *both* questions answered, for (3), by the first part of (i) above. The answers are supplied by the first part of the following variant of (i).

(i bis) *Both contingencies* (α), (β), *described in* (i) *under the assumption* (3), *can occur if* (3) *is replaced by* (9).

On the other hand, in the particular case (9) *of* (6), *it is possible to refine* (8) *to*

$$x(t^*) = o(1) \text{ as } t^* \to \infty, \tag{8 bis}$$

if the unbounded, open t^*-set (7), *which is independent of the pair of integration constants determining a solution $x(t)$ of* (1), *is suitably chosen.*

The second part of (i bis) may be interpreted as what can be saved from (10) if the proviso italicized before (10) is omitted.

Proof of (7), (8)-(8 bis). Let $N(t)$ denote the number of zeros possessed by a solution of an arbitrary differential equation (1) on the interval $[0, t]$. It is known (cf., e. g., [6]) that, if $x_1(t)$ and $x_2(t)$ are two linearly independent solutions of (1), and if their Wronskian, which is always a non-vanishing constant, is normalized to be 1 (or -1), then

$$\pi N(t) = \int_0^t du/r^2(u) + O(1) \text{ as } t \to \infty, \tag{11}$$

where

$$r = (x_1^2 + x_2^2)^{\frac{1}{2}} > 0 \tag{12}$$

(in fact, $x_1 = x_1(t)$ and $x_2 = x_2(t)$ cannot vanish simultaneously). Needless to say, the step-function $N(t)$, as defined *before* (11), depends on the choice of the particular solution, say $x(t)$, whose zeros $N(t)$ enumerates. On the other hand, it follows from Sturm's separation theorem that two N-functions belonging to two solutions of (1) cannot differ by more than 1 at any t, and so the indeterminacy of $N(t)$ is absorbed by the $O(1)$ in (11).

All of this is independent of the assumption (6). It will be shown that, according as (6) or its refinement (9) is assumed,

$$\liminf r(t) < \infty \text{ or } \liminf r(t) = 0, \qquad (t \to \infty), \tag{13}$$

respectively. This, when combined with the continuity of the function $r(t)$, will prove the respective assertions of (i), (i bis) concerning the existence of open, unbounded sets (7) on which (8), (8 bis) hold. In order to see this, it is sufficient to compare (13) with (12) and to observe that, since $x_1(t)$ and $x_2(t)$ are linearly independent, every solution of (1) is of the form

$$x(t) = c_1x_1(t) + c_2x_2(t), \tag{14}$$

where c_1 and c_2 are independent of t.

For large t, the linear oscillator $x'' + \omega^2x = 0$ is a "Sturmian minorant" of (1) for some or for every choice of the positive constant ω, according as (6) or (9) is assumed for (1). Since the asymptotic density of the zeros of the solutions of the linear oscillator is proportional to ω, it follows that, if $N(t)$ refers to (1),

$$\liminf N(t)/t > 0 \text{ or } N(t)/t \to \infty \qquad (t \to \infty),$$

according as (6) or (9) is assumed. Finally, it is seen from (11) that the last formula line implies the truth of the respective assertions (13).

The remaining four assertions, viz., (α) and (β) in (i) and (i bis), will now be proved in order.

Proof of (α) *in* (i). Corresponding to every $a > 0$, it is possible to choose an $f(t)$ satisfying (3) and having the property that some solution of (1) becomes $O(t^{-a})$ as $t \to \infty$; cf. [20], p. 269. In particular, (3) is compatible with the existence of a solution satisfying (10). But if $x(t)$ is a solution satisfying (10), then any solution linearly independent of $x(t)$ must violate (5); cf. [18], p. 397.

Proof of (β) *in* (i). This contingency, too, is between the lines of [18], the situation being as follows:

(A) If $f(t)$ is subject to (3), and if a solution $x = x^*(t)$ of (1) is such as to satisfy

$$\liminf a^*_n = 0, \qquad (n \to \infty),$$

where $a^*_1, a^*_2, \cdots$ is the sequence of the relative maxima of $|x^*(t)|$, then every solution $x(t)$ linearly independent of this $x^*(t)$ must violate (5). In fact, if the assumption made by the last formula line is strengthened to $a^*_n \to 0$ as $n \to \infty$, then, since the latter assumption is equivalent to the case $x = x^*$ of (10), the violation of (5) by every $x(t) \neq \text{const.}\, x^*(t)$ becomes the assertion used at the end of the above Proof of (α) in (i). On the other hand, a glance at the proof ([18], p. 397) of this assertion shows that what is actually used there is just the assumption made in the last formula line, rather than its strengthened form, which is (10).

(B) According to [18], p. 396, a function $f(t)$ satisfying (3) can be chosen as follows: *Some* solution $x(t)$ of (1) satisfies *both*

$$\liminf a_n = 0 \text{ and } \limsup a_n = \infty \qquad (n \rightarrow \infty),$$

where $a_1, a_2, \cdots$ denotes the sequence of the relative maxima of $|x(t)|$.

If (A) is compared with (B), it is seen that there exists an $f(t)$ satisfying (3) and belonging to the contingency (β) of (i).

Proof of (α) *and* (β) *in* (i bis). The proof depends on an adaptation of a known principle of construction (cf. [15]; also [13]).

Choose a sequence of positive numbers $\lambda_1, \lambda_2, \cdots$ satisfying

$$(15) \qquad \lim_{k\rightarrow\infty} \lambda_k = \infty \text{ and } \sum_{k=1}^{\infty} 1/\lambda_k = \infty,$$

put $\beta_k = \frac{1}{2}\pi \sum_{j=1}^{k-1} 1/\lambda_j$, where $\beta_1 = 0$, and define $f(t)$ by placing

$$f(t) = \lambda_k^2 \text{ when } \beta_k \leqq t < \beta_k \qquad (k = 1, 2, \cdots).$$

Thus $f(t)$ is defined for $0 \leqq t < \infty$ by virtue of (15). This $f(t)$, instead of being continuous throughout, is a step-function but, as far as the purposes at hand go, the jumps of $f(t)$ can be smoothed out by "small" local alterations of the graph $f = f(t)$. Clearly, (9) is satisfied.

Let $d_1, d_2, \cdots$ and $e_1, e_2, \cdots$ denote the sequences of numbers

$$(16) \qquad d_k = \prod_{j=1}^{k-1} \lambda_{2j}/\lambda_{2j+1} \text{ and } e_k = \prod_{j=1}^{k-1} \lambda_{2j+1}/\lambda_{2j+2};$$

so that

$$(17) \qquad e_k = \lambda_2/(\lambda_{2k} d_k).$$

Put

$$x_1(t) = (-1)^k d_k \cos \lambda_{2k}(t - \beta_{2k}) \text{ or } x_1(t) = (-1)^{k+1} d_{k+1} \sin \lambda_{2k+1}(t - \beta_{2k+1})$$

according as

$$\beta_{2k} \leqq t < \beta_{2k+1} \text{ or } \beta_{2k+1} \leqq t < \beta_{2k+2}$$

(so that $x_1(t)$ is defined for $\beta_2 \leqq t < \infty$), and put

$$x_2(t) = (-1)^k e_k \cos \lambda_{2k+1}(t - \beta_{2k+1}) \text{ or } x_2(t) = (-1)^{k+1} e_{k+1} \sin \lambda_{2k+2}(t - \beta_{2k+2})$$

according as

$$\beta_{2k+1} \leqq t < \beta_{2k+2} \text{ or } \beta_{2k+2} \leqq k < \beta_{2k+3}$$

(so that $x_2(t)$ is defined for $\beta_3 \leqq t < \infty$). Then, since $f(t)$ is the constant λ_k^2 when $\beta_k \leqq t < \beta_{k+1}$, both $x = x_1(t)$ and $x = x_2(t)$ are solutions of (1) on each of the successive intervals $\beta_k < t < \beta_{k+1}$. In addition, the constants are adjusted so as to assure the existence of continuous *first* derivatives $x'_1(t)$, $x'_2(t)$ throughout (i.e., at the points $t = \beta_k$ also). This is seen from the definition of the β_k's.

Ad (α). Choose

$$\lambda_{2j} = j^2 \text{ and } \lambda_{2j+1} = j \qquad (j = 1, 2, \cdots);$$

so that (15) is satisfied. Furthermore,

$$d_k = (k-1)! \text{ and } e_k = 1/[k-1)!k^2],$$

by (17). Hence, by the above definitions of $x_1(t)$ and $x_2(t)$,

$$x_1(t) \neq O(1) \text{ but } x_2(t) = o(1), \text{ as } t \rightarrow \infty.$$

In view of (14), this means that a solution $x(t)$ is bounded, as $t \rightarrow \infty$, if and only if $x(t) = \text{const.}\, x_2(t)$. Consequently, the present $f(t)$ is of the kind required by (α) in (i bis).

Ad (β). Choose a sequence $\lambda_1, \lambda_2, \cdots$ satisfying (15) and having the property that

$$(18) \qquad \limsup d_k = \infty \text{ and } \limsup e_k = \infty \qquad (k \rightarrow \infty)$$

hold for the numbers defined by (16). [Such sequences $\lambda_1, \lambda_2, \cdots$ exist; one results by placing

$$\lambda_{2j} = j^2 \text{ and } \lambda_{2j+1} = j \text{ if } k_{2n} \leqq j < k_{2n+1}$$

and

$$\lambda_{2j} = j \text{ and } \lambda_{2j+1} = j^2 \text{ if } k_{2n+1} \leqq j < k_{2n+2},$$

where $1 = k_0 < k_1 < \cdots$ is a suitable sequence of integers.] Then, if the pair of alternative formulae defining $x_1(t)$ and $x_2(t)$ is compared with (16), (17) and (18), it is readily verified that a linear superposition (14) of $x_1(t)$ and $x_2(t)$ is bounded as $t \rightarrow \infty$ only if both constants c_1, c_2 of (14) vanish. This means that (1) has no bounded solution ($\neq 0$), as required by (β) in (i bis).

The Non-Oscillatory Case.

If a solution (and therefore, by Sturm's separation theorem, every solution) of (1) has an infinity of zeros on the half-line (2), then (1) is called oscillatory, and otherwise non-oscillatory. Since the zeros of a solution ($\not\equiv 0$) cannot cluster at a finite $t = t_0$, it is seen from (11) and (12) that (1) is non-oscillatory if and only if some (and/or every) pair of linearly independent solutions, $x = x_1(t)$ and $x = x_2(t)$, of (1) satisfies

$$\int_0^\infty dt/\{x_1^2(t) + x_2^2(t)\} < \infty. \tag{19}$$

Further information is contained in the following statements:

(ii) *If* (1) *is non-oscillatory, then, as* $t \to \infty$,

- (a) *some solution must fail to be* $O(t^{\frac{1}{2}})$;
- (b) *every solution can* (*but need not*) *be* $O(t^{\frac{1}{2}} \log t)$;
- (c) *every solution can* (*but need not*) *satisfy* $t^{\frac{1}{2}} = O(|x(t)|)$;
- (d) *some solutions must satisfy*

$$\int^\infty dt/x^2(t) < \infty. \tag{20}$$

(a) is a consequence of (19); cf. [23]. The negative (parenthetical) remarks of (b) and (c) follow from either of the examples $f(t) \equiv -1$, $f(t) \equiv 0$, since the general solution of (1) then is $x(t) = c_1 e^t + c_2 e^{-t}$, $x(t) = c_1 + c_2 t$, respectively. The positive assertions of (b) and (c) result by choosing $f(t) = (2t)^{-2}$ (for $1 \leqq t < \infty$) and observing that the general solution of (1) then is $x(t) = c_1 t^{\frac{1}{2}} + c_2 t^{\frac{1}{2}} \log t$.

The remaining assertion, (d), was proved in [2]. Needless to say, (d) is a refinement of (19). It is understood that the lower limit of integration in (20) is meant to be large enough, viz., greater than the last zero of the particular solution in question.

Roughly speaking, the above comments could be interpreted by saying that, if (1) is non-oscillatory, there exists a solution which, as $t \to \infty$, is "small" (in this regard, cf. (a)-(b) in (v) below), whereas the other solutions, those linearly independent of the "small" one, are "large" (in some sense).

Actually, the assertions of (ii) contain about everything that can be said of the solutions of an arbitrary non-oscillatory differential equation (1). This can be seen by constructions similar to the one which will lead to the following contingency:

(iii) *A differential equation* (1) *can be non-oscillatory and nevertheless such that there exists an unbounded, open set* (7) *which is independent of the integration constants and has the following property: As t tends to ∞ on* (7), *every solution of* (1) *tends to* 0.

Much less than this restricted $o(1)$-situation is prevented by (a) in (ii) if t, instead of being restricted to a suitable set (7), tends to ∞ without a restriction.

In order to prove (iii), choose two monotone sequences of positive numbers, say $k_0, k_1, \cdots$ and $\lambda_0, \lambda_1, \cdots$, satisfying

$$\lim_{n\to\infty} k_n = \infty \text{ and } \sum_{t<n}^{\infty} k_n{}^2/\lambda_n = O(e^{-2t}), \qquad t \to \infty \tag{21}$$

(such pairs of sequences exist, as seen by choosing $k_n = n$ and $\lambda_n = n^4 e^{2n}$ if $n > 0$). For the purposes of (22) below, choose λ_0 greater than π.

In terms of the numbers k_n, λ_n, define on the half-line (2) a function $x_1(t)$ by placing

$$\begin{aligned} 1/x_1(t) &= e^{-t} \text{ if } n + \pi/\lambda_n < t < n + 1, \\ 1/x_1(t) &= e^{-t} + k_n \sin^4 \lambda_n (t - n) \text{ if } n \leqq t \leqq n + \pi/\lambda_n, \end{aligned} \tag{22}$$

where $n = 0, 1, 2, \cdots$. It is seen from (22) that $x_1(t)$ has a continuous second derivative on (2) and that

$$x_1(t) > 0 \text{ on } (2). \tag{23}$$

Hence, if $f(t)$ is defined by

$$f(t) = -x''_1(t)/x_1(t), \tag{24}$$

then $f(t)$ is continuous on (2). The corresponding (1) is satisfied by $x = x_1(t)$; cf. (24). It follows therefore from (23) that (1) is non-oscillatory.

It is readily seen from (21), (22) and (23) that (20) is satisfied by $x = x_1(t)$. This fact, when combined with (23), means that

$$x_2(t) = x_1(t) \int_t^\infty du/x_1{}^2(u) \tag{25}$$

defines on (2) a function, $x_2(t)$. Since (1) is satisfied by $x = x_1(t)$, two differentiations of (25) show that (1) is satisfied by $x = x_2(t)$ also. Furthermore, (25) alone implies that

$$x_1(t)x'_2(t) - x_2(t)x'_1(t) = -1, \tag{26}$$

which in turn implies that the solutions $x_1(t)$, $x_2(t)$ are linearly independent.

According to (22),

$$\int_t^\infty du/x_1^2(u) \leqq 2\int_t^\infty e^{-2u}du + O\Big(\sum_{t<n}^{\infty} k_n^2/\lambda_n\Big)$$

as $t \to \infty$. Both terms on the right of this inequality are $O(e^{-2t})$, by the second of the relations (21). Hence, (25) and the last formula line imply that $x_2(t) = x_1(t)O(e^{-2t})$. On the other hand, $x_1(t) = O(e^t)$, by (22). Consequently, $x_2(t) = O(e^t)O(e^{-2t})$ and so, in particular,

$$\lim x_2(t) = 0 \qquad (t \to \infty).$$

Finally, it is seen from (22) and from the first of the relations (21) that

$$\liminf x_1(t) = 0 \qquad (t \to \infty).$$

Since $x_1(t)$ and $x_2(t)$ are continuous, it is seen from (23) and (14) that the last two formula lines complete the proof of (iii).

The general properties of an arbitrary non-oscillatory differential equation (1), which are collected in (ii), can be completed by the following comparison theorem:

(iv) *Let* $f(t)$, $g(t)$ *be continuous functions satisfying*

$$f(t) \geqq g(t) \tag{27}$$

on (2), *and suppose that* (1) *is non-oscillatory. Then, if* $x = x(t)$ *is any solution of the differential equation* (1), *the differential equation*

$$y'' + g(t)y = 0 \tag{28}$$

(*is non-oscillatory and*) *has some solution* $y = y(t) = y_x(t)$ *satisfying*

$$y(t) = O(|x(t)|) \text{ as } t \to \infty. \tag{29}$$

The parenthetical assertion of this comparison theorem follows, by (27), from Sturm's comparison theorem.

In order to prove (iv), let $x(t)$ be a given solution of (1). Since (1) is non-oscillatory, it can be assumed that

$$x(t) > 0 \text{ for } t_0 < t < \infty, \tag{30}$$

if t_0 is large enough. Let t be on the half-line $t_0 < t < \infty$. Then, if $y(t)$ is any function, there exists a unique function $z(t) = z_x^y(t)$ satisfying

$$y(t) = z(t)x(t). \tag{31}$$

This makes applicable the method of the variation of constants, as follows:

Substitution of (31) into (28) gives

$$z''x + 2z'x' - zfx + gzx = 0,$$

since $x'' = -fx$, by (1). If this differential equation for $z = z(t)$ is multiplied by the given function $x = x(t)$ (which, according to (30), is an admissible operation), what results can be written in the form

$$(p(t)z')' - q(t)z = 0, \tag{32}$$

where $p = x^2$ and $q = (f - g)x^2$; so that, by (30) and (27),

$$p(t) > 0 \text{ and } q(t) \geqq 0. \tag{33}$$

It is now seen from (31) that (iv) is contained in the following lemma:

(iv*) *If $p(t)$ and $q(t)$ are continuous functions satisfying* (33) *on* (2), *then* (32) *has a solution $z(t)$ which is $O(1)$ as $t \rightarrow \infty$.*

(iv*) is known; it can be deduced (cf. [8]) as a corollary of a theorem of Kneser [11], dealing with the particular case $p(t) \equiv 1$.

An illustration of (iv bis) can be formulated as follows:

If $f(t) \leqq -a/t^2$ holds, for some $a = \text{const.} > 0$, as $t \rightarrow \infty$, then (1) *has a "small" solution of the form $O(t^{-b})$, where $b = b(a) > 0$:*

$$2b = (1 + 4a)^{\frac{1}{2}} - 1. \tag{34}$$

In order to see this, it is sufficient to choose $g(t)$, $f(t)$ in (iv bis) to be $f(t)$, $-a/t^2$, respectively (if $t \geqq 1$), and to observe that, if b is defined by (34), the case $f(t) = -a/t^2$ of (1) is satisfied by $x(t) = t^{-b}$.

The Case of a Non-Oscillatory λ-Region.

Some of the assertions and negations of (ii), (iii), (iv) can be refined, and the construction proving (iii) adapted to more delicate situations, if the non-oscillatory character of (1) is so manifest that

$$x'' + (f(t) + \lambda)x = 0 \tag{35}$$

is non-oscillatory for some, sufficiently small, $\lambda = \text{const.} > 0$. This is a strengthening of the assumption that (1) is non-oscillatory. In fact, the parenthetical remark in (iv) implies that, if (35) is non-oscillatory for some λ, then (35) is non-oscillatory for every λ which is smaller than the given λ.

Accordingly, there belongs to every $f(t)$ a λ^* having the property that (35) is non-oscillatory if

$$-\infty < \lambda < \lambda^*, \tag{36}$$

and oscillatory if $\lambda^* < \lambda < \infty$, where it is understood that both limiting cases $\lambda^* = \pm\infty$ are included. If λ^* is called the parabolic point of the λ-axis, and if $\lambda^* \neq \pm\infty$, then it depends on $f(t)$ whether (35) is oscillatory or non-oscillatory for $\lambda = \lambda^*$. Even in the latter case, some of the following considerations do not apply at $\lambda = \lambda^*$.

(v) *Suppose that*

$$-\infty < \lambda^* \leqq \infty, \tag{37}$$

where λ^* *denotes the parabolic point of* (35). *Then*

(a) *there belongs to every point of the region* (36) *a solution* $x(t) = x_\lambda(t) \not\equiv 0$ *of* (35) *which is of class* (L^2) *on* (2);

(b) *for every* $\lambda < \lambda^*$, *this* $x_\lambda(t)$ *is uniquely determined to a constant factor*;

(c) *although, by* (a),

$$0 < \int_0^\infty x_\lambda^2(t)\,dt < \infty, \qquad -\infty < \lambda < \lambda^*, \tag{38}$$

it is not in general true that

$$x_\lambda(t) = O(1), \qquad t \to \infty, \tag{39}$$

where $\lambda < \lambda^*$;

(d) *if* (39) *holds for some* $\lambda = \mu < \lambda^*$, *then it holds for every* $\lambda < \mu$;

(e) *if*

$$x_\lambda(t) = o(1), \qquad t \rightarrow \infty, \tag{40}$$

holds for some $\lambda = \mu < \lambda^*$, *then it holds for every* $\lambda < \mu$;

(f) *if* (39) *holds for some* $\lambda = \mu < \lambda^*$, *then* (40) *need not hold for every* $\lambda < \mu$; *however,*

(g) *if* (39) *holds but* (40) *does not hold for a pair of values* $\lambda = \lambda_1, \lambda_2$, *where* $\lambda_1 < \lambda_2 < \lambda^*$, *then* $\lambda^* = \infty$, *and*

$$\lim_{t\rightarrow\infty} x_\lambda(t)/x_\mu(t) \text{ exists and is not } 0 \tag{41}$$

for $-\infty < \lambda < \mu < \infty$.

(a) was proved in [2] (by using (d) in (ii) above).

(b) can be concluded as follows: If (b) is assumed to be false, then, for some value of λ, two linearly independent, hence all, solutions of (35) are of class (L^2) on (2); so that (35) is of *Grenzpunkt* type. But this is impossible if (35) is non-oscillatory for some λ (cf. [2] and, for a sharper result, [6]). This proves (b).

(d) and (e) are clear from (iv bis). In order to prove (c), (f) and (g), the following lemma will first be established:

LEMMA. *If* (1) *is non-oscillatory and has a solution* $x = x(t)$ *satisfying*

$$\int_T^\infty x^2(t)\Big(\int_T^t du/x^2(u)\Big)dt < \infty \qquad (\text{for some } T > 0), \tag{42}$$

then (35) *possesses, for every* λ, *a pair of (linearly independent) solutions,* $x_1 = x_1(t, \lambda)$ *and* $x_2 = x_2(t, \lambda)$, *satisfying*

$$x_1(t, \lambda) \sim x(t) \text{ and } x_2(t, \lambda) \sim x(t) \int_T^t du/x^2(u) \tag{43}$$

as $t \rightarrow \infty$; *conversely, if the first relation in* (43) *holds for some* $\lambda \neq 0$ *and for some* (L^2)*-solution* $x = x(t)$ *of* (1) *on* (2), *then the latter satisfies* (42).

Proof of the Lemma. In (28), let $g(t) = f(t) + \lambda$ (so that (28) and (35) are equivalent). The relation (27) will not hold unless $\lambda \leqq 0$; however, this is immaterial in what follows.

The variation of constants, (31), transforms (28) into (32), where

$$p(t) = x^2(t) \text{ and } q(t) = -\lambda x^2(t).$$

If the new independent variable

$$s = s(t) = \int_T^t du/p(u) = \int_T^t du/x^2(u) \tag{44}$$

is introduced, then (32) becomes

$$\ddot{z} - Q(s)z = 0 \qquad (\cdot = d/ds), \tag{45}$$

where

$$Q(s) = p(t(s))q(t(s)) = -\lambda x^4(t(s)) \tag{45 bis}$$

and $t = t(s)$ is the inverse function of (44).

If $x(t)$ is of class (L^2), which is certainly the case if (42) holds, then $s(t) \rightarrow \infty$ as $t \rightarrow \infty$. In fact, since the harmonic mean of a positive function does not exceed its arithmetical mean,

$$\int_T^t du/x^2(u) \geqq (t-T)^2 / \int_T^t x^2(u)\,du.$$

Hence, the truth of the assertion, $s(\infty) = \infty$, is seen from (44).

Since the function (45 bis) does not change sign, it follows that (45) possesses a solution $z = z(s)$ satisfying

$$z(s) \rightarrow 1 \text{ as } s \rightarrow \infty \tag{46}$$

if and only if

$$\int^{\infty} s\,|\,Q(s)\,|\,ds < \infty. \tag{47}$$

In fact, whether $Q(s)$ does or does not change its sign, a result of Bôcher ([1], § 4) supplies the sufficiency of (47) for the existence of a solution $z = z(s)$ of (45) satisfying (46). It is also known (cf. [16], pp. 40-42 or [8], Appendix, where $Q \leqq 0$, and [5], pp. 533-534, where $Q \geqq 0$) that this sufficient condition is necessary as well if Q does not change sign.

If $\lambda \neq 0$, then (44) and (45 bis) show that (47) is equivalent to (42). It therefore follows from (31) that, if $x(t)$ is of class $L^2(0, \infty)$, then (42) is necessary and sufficient in order that (35) has, for some $\lambda \neq 0$ (and/or for every λ), a solution $x = x_1(t, \lambda)$ satisfying the first of the relations in (43). Finally, the second of the relations in (43) is a consequence of the first, since

$$x = y(t) \int_T^t du/y^2(u)$$

is a solution of (35) whenever y is a non-vanishing solution for $t \geqq T$.

This completes the proof of the Lemma.

Proof of (g) *in* (v). Without loss of generality, assume that $\lambda_2 = 0$, hence $0 < \lambda^* \leqq \infty$, and that $x = x(t)$ is an (L^2)-solution of (1) on (2) and is $O(1)$, but is not $o(1)$, as $t \to \infty$. Let $\lambda < 0$ and $g(t) = f(t) + \lambda$. Apply the variation of constants, (31), which reduces (28) and/or (35) to (32). Then (33) holds (since $\lambda < 0$). Hence, (iv*) assures the existence of a bounded solution $z = z(t)$ of (32). Actually, this bounded solution can be chosen in such a way that $z(t) > 0$ and $z'(t) \leqq 0$ (cf. [8]); in particular,

$$\lim_{t \to \infty} z(t) \geqq 0 \text{ exists.} \tag{48}$$

Since $x(t)$ is $O(1)$ but is not $o(1)$, it is clear from (31) and (48) that an (L^2)-solution of (28) and/or of (35) is $o(1)$ if and only if the sign of equality holds in (48).

Accordingly, the assumption of (g), namely, that there exists a $\lambda_1 < \lambda_2 = 0$ violating the case $\lambda = \lambda_1$ of (40), implies that the limit (48) is not 0. Consequently, if $z(t)$ is multiplied by a suitable constant, the limit in (48) can be chosen to be 1. Then (31) shows that the first relation in (43) holds if $x_1(t, \lambda) = x_\lambda(t)$ and $\lambda = \lambda_1 \neq 0$. In view of the Lemma, this completes the proof of (g) in (v).

Proof of (c) *and* (f) *in* (v). On (2), let $\phi = \phi(t)$ be a continuous decreasing function satisfying

$$\int_0^\infty \phi^2(t)\,dt < \infty; \tag{49}$$

so that $\phi(t) > 0$, and $\phi(t) \to 0$ as $t \to \infty$.

Let $k_1, k_2, \cdots$ and $\lambda_1, \lambda_2, \cdots$ be two sequences of positive numbers satisfying $k_n \geqq 1$, $\lambda_n > \pi$ and

$$\sum_{n=1}^\infty (k_n^2/\lambda_n) \int_0^{n+1} dt/\phi^2(t) < \infty \qquad (\text{hence } \sum_{n=1}^\infty k_n^2/\lambda_n < \infty). \tag{50}$$

Define on (2) a function $x(t)$ by placing

$$(51)\qquad \begin{aligned} &x(t)=\phi(t) \text{ if either } 0\leqq t<1 \text{ or } n+\pi/\lambda_n<t<n+1,\\ &x(t)=\phi(t)+k_n\sin^4\lambda_n(t-n) \text{ if } n\leqq t\leqq n+\pi/\lambda_n,\end{aligned}$$

where $n=1,2,\cdots$. Clearly, $x(t)>0$ on (2).

If $\phi(t)$ has a continuous second derivative on (2), then the same is true of $x(t)$, by (51). Hence, the case $x=x_1$ of (24) defines a continuous $f(t)$ on (2). For this $f(t)$, the function (51) is a solution of (1), by (24), and is of class (L^2) on (2), by (49) and (50).

It will be shown that if $\phi(t)$ is suitably chosen, then (42) holds for the $x=x(t)$ defined by (51).

First, $x(t)\geqq\phi(t)$, by (51). Hence

$$\int_0^t du/x^2(u)\leqq\int_0^t du/\phi^2(u).$$

It follows, therefore, from (51) and from Minkowski's inequality, that (42) is satisfied if

$$(52)\qquad \int_0^\infty \phi^2(t)\left(\int_0^t du/\phi^2(u)\right)dt<\infty$$

and

$$(53)\qquad \sum_{n=1}^\infty\int_n^{n+\pi/\lambda_n} k_n^2\left(\int_0^t du/\phi^2(u)\right)dt<\infty.$$

But (53) is a consequence of (50). In order to satisfy (52), it is sufficient to choose $\phi(t)=t^{-1}\exp(-t^3)$ when $1\leqq t<\infty$. For then

$$\int_1^t du/\phi^2(u)=\int_1^t u^2\exp(2u^3)\,du=O(1)\exp(2t^3),$$

hence

$$\int_1^\infty \phi^2(t)\left(\int_0^t du/\phi^2(u)\right)dt=O(1)\int_1^\infty t^{-2}dt<\infty.$$

Accordingly, all of the assumptions of the first part of the Lemma are fulfilled. Hence, (35) possesses, for every λ, a pair of solutions $x_1(t,\lambda)$, $x_2(t,\lambda)$ satisfying (43).

In order to complete the proof of (c) and (f) in (v), it is sufficient to choose $k_n=n$ and $k_n=1$, respectively. In fact, it is clear from (51) that

$x(t)$ is not $O(1)$ in the first case, and that $x(t)$ is $O(1)$ but is not $o(1)$ in the second case .Finally, it is seen from (43) that $x(t)$ can here be replaced by $x_\lambda(t) = x_1(t, \lambda)$ in both cases, where $-\infty < \lambda < \infty$.

The Parabolic Case.

Before (v), the parabolic case, $\lambda = \lambda^*$, was defined as follows: (35) is oscillatory or non-oscillatory according as $\lambda > \lambda^*$ or $\lambda < \lambda^*$, where $\lambda^* \neq \pm \infty$. It will now be shown that, in contrast to the positive assertions of (v), which assume (36), anything can happen when $\lambda = \lambda^*$.

(vi) *Suppose that* $\lambda^* \neq \pm \infty$, *where* λ^* *denotes the parabolic point of* (35). *Then the case* $\lambda = \lambda^*$ *of* (35) *can be*

(a) oscillatory, (b) non-oscillatory

and it can have on (2)

(α) an (L^2)-solution (β) no (L^2)-solution.

In fact, all four possibilities

$$(a\alpha), \quad (a\beta), \quad (b\alpha), \quad (b\beta) \tag{54}$$

can be realized. What cannot occur is that the case $\lambda = \lambda^*$ *of* (35) *be such as to have two, linearly independent,* (L^2)*-solutions on* (2).

First, whether $\lambda^* \neq \pm \infty$ is or is not assumed, (35) cannot have two, linearly independent, (L^2)-solutions on (2) for any λ unless the same is true for every λ. This is Weyl's theorem (cf. [17], p. 238) on which his alternative of "*Grenzpunkt* case or *Grenzkreis* case" depends (for a sharper result, establishing an explicit asymptotic connection between the solutions belonging to two arbitrary λ-values in the *Grenzkreis* case, cf. [20], pp. 266-267). Accordingly, the last assertion of (vi) follows from assertion (b) of (v).

Next, since $x'' + \lambda x = 0$ is oscillatory or non-oscillatory according as $\lambda > 0$ or $\lambda < 0$, it follows from Sturm's comparison theorem that if

$$f(t) \to 0 \text{ as } t \to \infty, \tag{55}$$

then (35) is oscillatory or non-oscillatory according as $\lambda > 0$ or $\lambda < 0$. In other words, (55) is sufficient for $\lambda^* = 0$. In particular, if $f(t) = Ct^{-2}$, where C is any real constant (and, for instance, $1 \leqq t < \infty$, whilst $f(t) = C$ when $0 \leqq t < 1$), then $\lambda^* = 0$; so that the case $\lambda = \lambda^*$ of (35) becomes

$$x'' + Ct^{-2}x = 0. \tag{56}$$

It is easy to see that, by suitable values of C, each but the first of the four cases (54) can be realized by (56); cf. [10], pp. 415-416.

In fact, (56) is satisfied by $x(t) = t^c$ if c is a root of a quadratic equation,

$$2c = 1 \pm (1 - 4C)^{\frac{1}{2}}. \tag{34 bis}$$

If $\frac{1}{4} < C < \infty$, the two roots c are distinct, complex and have the real part $\frac{1}{2}$; so that (56) is oscillatory and, since $t^{\frac{1}{2}}$ is not of class (L^2) on (2), the second of the possibilities (54) takes place. If $-\infty < C < \frac{1}{4}$, the two roots are real and distinct; furthermore, the smaller c is less than $-\frac{1}{2}$ if and only if $-\infty < C < -\frac{3}{4}$. Since $x(t) = t^c$ is a solution, it follows that (56) realizes the third or the fourth of the possibilities (54) according as $-\infty < C < -\frac{3}{4}$ or $-\frac{3}{4} \leqq C < \frac{1}{4}$.

The remaining possibility, that which is the first in (54), is more delicate. Its occurrence will now be constructed by an adaptation of the "Riccati" method used in [20], p. 268 (where the λ-value, instead of being at a sharply defined parabolic point, is well within an "elliptic" range); cf. also [22].

Proof of $(a\alpha)$ *in* (vi). Let (1) be replaced by $1 \leqq t < \infty$, put

$$x(t) = r(t) \cos \log t, \qquad (1 \leqq t < \infty), \tag{57}$$

where, for the present, the function $r(t)$ is undetermined, and substitute (57) into (1). This gives

$$f(t) = t^{-2} - r''(t)/r(t) + \{2r'(t)/r(t) - t^{-1}\}\{t^{-1} \tan \log t\}, \tag{58}$$

(58) being computed from (57) and (1) as $f = -x''/x$.

In order to make the function (58) continuous on the half-line $1 \leqq t < \infty$, the infinities of the factor $\tan \log t$, which cluster at $t = \infty$, must be neutralized by an appropriate choice of the first factor of the last term of (58). Since that first factor is the derivative of

$$2 \log r(t) - \log t = \log r^2(t)/t,$$

such a neutralization results by choosing the logarithmic derivative of $r^2(t)/t$ to be of the form $\chi(t) \cos^2 \log t$, where $\chi(t)$ is an arbitrary continuous function on the half-line $1 \leqq t < \infty$. If $\chi(t) = -2\gamma t^{-1}$, where γ is a constant (which will be chosen below), this gives

$$r(t) = t^{\frac{1}{2}} \exp\left(-\gamma \int_1^t s^{-1} \cos^2 \log s \, ds\right). \tag{59}$$

If $r'(t)$ and $r''(t)$ are computed from (59), it is found that (58) reduces to

$$(60) \qquad 4t^2 f(t) = 5 - 4\gamma^2 \cos^4 (\log t) - 8\gamma \sin (2 \log t).$$

Clearly, (60) defines, for $1 \leqq t < \infty$, a continuous $f(t)$, and (57), with (59), represents a solution of (1) for this $f(t)$.

Since (55) is satisfied by (60), and since (55) is sufficient for $\lambda^* = 0$, the parabolic case of (35) becomes (1). Hence, the possibility ($a\alpha$), claimed in (vi), will be realized if it is shown that (1) now possesses a solution which has an infinity of zeros and is of class (L^2). But

$$\int_1^t s^{-1} \cos^2 \log s \, ds = \int_0^{\log t} \cos^2 s^* ds^* = \tfrac{1}{2} t^* + \tfrac{1}{4} \sin 2t^*,$$

where $s^* = \log s$, $t^* = \log t$. Hence, from (57) and (59),

$$x(t) = (\cos \log t) t^{\frac{1}{2} - \frac{1}{2}\gamma} \exp(- \tfrac{1}{4}\gamma \sin 2 \log t).$$

Consequently, $x(t)$ is of class (L^2) on the half-line $1 \leqq t < \infty$ if the constant γ, which thus far was arbitrary, is chosen so as to satisfy $\frac{1}{2} - \frac{1}{2}\gamma < -\frac{1}{2}$. Since the last formula line also shows that $x(t)$ acquires an infinity of zeros as $t \rightarrow \infty$, the proof is complete.

Since the oscillating function (60) is $O(1)$ as $t \rightarrow \infty$, the $f(t)$ defined by (58) is of the same order, t^{-2}, as the coefficient function of (56). On the other hand, whilst the trivial family (56) sufficed for the realization of all but the first of the four possibilities (54), the delicate construction leading to (58) was needed for the realization of the first of these possibilities. This situation is explained by the following criterion:

(vi bis) *Case* ($a\alpha$) *of* (vi) *cannot occur if* $f(t)$ *is monotone for large* t.

Clearly, it is sufficient to prove this under the assumption that $f(t)$ is monotone throughout, i. e., on (2).

First, it is clear from Sturm's comparison theorem that, if $f(t)$ tends to a limit, $f(\infty)$, then (35) is non-oscillatory for every λ, oscillatory for every λ, or non-oscillatory for every $\lambda < f(\infty)$ and oscillatory for every $\lambda > f(\infty)$, according as $f(\infty) = -\infty$, $f(\infty) = \infty$, or $f(\infty) \neq \pm \infty$. Hence, $\lambda^* = f(\infty)$ holds in all three cases. Since (vi) assumes that $\lambda^* \neq \pm \infty$, and since the constant λ^* can be subtracted from $f(t)$ in (35) if the origin of the λ-axis is translated by λ^*, it follows that (vi bis) need be proved only under the assumption $f(\infty) = 0$, and that the case $\lambda = \lambda^*$ of (35) then becomes (1). In addition, $f(t)$ cannot change sign, since $f(t)$ is monotone. If $f(t) \leqq 0$, then, according to Sturm, (1) is in case (b) of (vi), and so (vi bis) does not claim anything. Let therefore $f(t) \geqq 0$. Then $df(t) \leqq 0$, since $f(t)$ is monotone and $f(0) = 0$.

Suppose that (1) is in case (α) of (vi). Then, since $f(t) = O(1)$ as $t \rightarrow \infty$, the (L^2)-solution in question is subject to

$$x(t) \rightarrow 0 \text{ as } t \rightarrow \infty;$$

cf. [19], p. 8. Suppose, in addition, that (1) is in case of (a) of (vi). Then, if $t_1 < t_2 < \cdots$, where $t_n \rightarrow \infty$ as $n \rightarrow \infty$, denote the zeros of $x(t)$, and if the maximum of $|x(t)|$ on the t-interval (t_{n-1}, t_n) is $|x(t_n{}^*)|$,

$$|x(t_1{}^*)| \leqq |x(t_2{}^*)| \leqq \cdots, \text{ where } t_n{}^* \rightarrow \infty \text{ as } n \rightarrow \infty.$$

In fact, these inequalities follow from the assumption $df(t) \leqq 0$; cf., e. g., [22].

Since the last two formula lines are contradictory, it follows that (a) is incompatible with (α). This proves (vi bis).

Remark. It is seen from the relevant part of the proof of assertion (c) of (v) that, if $f(t)$ is wobbly enough as $t \rightarrow \infty$, the determination of the parabolic λ-value, λ^*, of (35) can become quite difficult. That the task must depend on "transcendental" means, can be illustrated by the analytical machinery (such as infinite determinants) which, in view of the following criterion, must be mobilized if the determination of λ^* is required in such a comparatively harmless case as

$$f(t) = f(t+1) \qquad (0 \leqq t < \infty). \tag{61}$$

If $f(t)$ is periodic on (2), then (35) has a parabolic λ-value, $\lambda = \lambda^* \neq \pm \infty$, which can be characterized as follows: λ^* is the least λ-value corresponding to which (35) possesses a solution having the same period as $f(t)$.

The proof will be omitted, since it is between the lines of the known description ([9], p. 280; [12]) of the distribution of λ-intervals of stability and instability in the case (61).

Spectral Formulations.

In what follows, the terminology will be the same as in the latter part of the Proof of (c) in (v) above.

($\mathcal{A}$) Suppose that

$$-\infty \leqq \limsup_{t\to\infty} f(t) < \infty. \tag{62}$$

Then (35) is the *Grenzpunkt* type ([17]; for a sharper result, cf. [6]). It was shown in [7] that, under the assumption (62),

(I) $\lambda = 0$ must be in the spectrum determined by (35) and a boundary condition, $ax(0) + bx'(0) = 0$, if (1) has at least one solution, $x = x(t)$, which satisfies this boundary condition and is $O(1)$ as $t \to \infty$; and that

(II) $\lambda = 0$ must be in the (invariant) *essential* spectrum of (35) if (35) has two, linearly independent, solutions which are $O(1)$ as $t \to \infty$.

It turns out that the respective *sufficient* conditions of (I) and (II), represented by the $O(1)$-assumptions, are not necessary as well (in either case); not even if (62) is strengthened to (3). This follows from the actual occurrence of both possibilities realized by (i) above. In fact, if $f(t)$ satisfies (3) and if the limit (3) is denoted by μ, then, as shown in [3], the essential spectrum of (35) consists of the half-line $-\mu \leqq \lambda < \infty$.

Remark. In the limiting case (9), where $\mu = \infty$, it is no longer true that (35) must be of *Grenzpunkt* type, and so the limiting case of the theorem quoted at the end of (A) ought to be formulated as follows: If $f(t)$ satisfies (9) *and* is such that (35) is not of *Grenzkreis* type, then the essential spectrum of (35) consists of the entire line, $-\infty < \lambda < \infty$. We did not prove that this limiting case of the theorem quoted is true. On the other hand, it is surely true in the other limiting case, in that belonging to $\mu = -\infty$. In fact, what is then claimed is that the essential spectrum consists of the points λ satisfying $-\mu \leqq \lambda$ if $\mu = -\infty$. But this is equivalent to the statement that the essential spectrum is vacuous (i. e., that the Green kernels belonging to (35) are completely continuous), if

$$f(t) \to -\infty \text{ as } t \to \infty. \tag{9 bis}$$

And this assertion is known to be true; it is contained, as a limiting case, in a theorem of Weyl, which will be quoted at the beginning of ($\mathcal{B}$) below (and which his proof establishes for the limiting case, $\mu = -\infty$, also).

($\mathcal{B}$) Suppose that (35) has a parabolic point, $\lambda^* \neq \pm\infty$, and choose, without loss of generality, λ^* to be 0. Thus, with reference to any boundary condition, $ax(0) + bx'(0) = 0$, the number of those points of the point spectrum which are to the left of λ^* is finite or infinite according as (35) is non-oscillatory or oscillatory at $\lambda = \lambda^*$. In either case, those points of the spectrum which are to the left of λ^* belong to the point spectrum only. According to Weyl [17], pp. 252-256, all of this can be concluded by an adaptation of a classical argument of Sturm.

As shown in [2], a proof of this theorem of Weyl can be adjusted so as to avoid assumption (62). What is actually needed in the Sturmian argument is the more general proviso that the differential equation be of *Grenzpunkt* type. But Weyl deals with the arbitrary self-adjoint case,

$$(p(t)x')' + (f(t) + \lambda) = 0, \tag{1 bis}$$

where $p(t)$ is any positive, continuous function on (2), rather than with the particular case (1), where $p(t) \equiv 1$. In the latter case, the proviso is not needed, since $\lambda^* > -\infty$ alone implies that (1) is of *Grenzpunkt* type. Cf. [6], where a sharper theorem is proved.

Due to the above criterion concerning that portion of the point spectrum which is to the left of λ^*, the $f(t)$ defined by (60) leads to a peculiar situation. In fact, since (55) is satisfied, $\lambda^* = 0$. But (57) is a solution of (1), i. e., of the case $\lambda = \lambda^*$ of (35). Since (57) has arbitrarily large zeros, it follows that those points of the spectrum which are to the left of 0 form an infinite sequence which tends to 0.

This holds for the point spectrum belonging to any boundary condition. Calculate $x(0)$, $x'(0)$ from (55) and (57), and consider that particular boundary condition which is satisfied by the resulting initial values $x(0)$, $x'(0)$. Then, since (59) makes (57) a function of class (L^2), the point spectrum will contain $\lambda = 0$. Accordingly, $\lambda = 0$ *is in the point spectrum.* At the same time, $\lambda = 0$ *is a cluster point of the point spectrum.*

In addition, $\lambda = 0$ *is in the continuous spectrum* also. In fact,

$$\int^{\infty} |f(t)|\, dt < \infty, \tag{63}$$

since $f(t) = O(t^{-2})$, by (60). But (63) is known to imply that every point of the closed half-line $0 \leqq \lambda < \infty$ is in the continuous spectrum of (35) and, incidentally, that no point of the open half-line $0 < \lambda < \infty$ is in the point spectrum. A simple proof results by observing that, if (63) is satisfied, then, according to Bôcher [1], two linearly independent solutions of (35) are of the form $x_1 = \cos\omega t + o(1)$, $x_2 = \sin\omega t + o(1)$ or $x_1 = e^{\omega t} + o(e^{\omega t})$, $x_2 = e^{-\omega t} + o(e^{-\omega t})$ according as $\lambda < 0$ or $\lambda > 0$, where $\omega = |\lambda|^{\frac{1}{2}}$ and $t \rightarrow \infty$ in both cases. On the other hand, in order that two linearly independent solutions of (35) be of the form $x_1 = 1 + o(1)$, $x_2 = t + o(t)$ in the parabolic case $\lambda = 0$, Bôcher requires

$$\text{(63 bis)} \qquad \int^{\infty} | tf(t) | \, dt < \infty,$$

rather than just (63). This explains why $\lambda = 0$ *can* be in a point spectrum belonging to the above example, $f(t)$ being just about t^{-2}; cf. (60).

(𝒷) Let $f(t)$ be any function for which (35) becomes of *Grenzpunkt* type, and let λ be any fixed value not contained in the essential spectrum (provided that there exists such a λ). Then (35) has a solution, $x = x_\lambda(t)$, which is of class (L^2) on (2) (and, to a constant factor, is uniquely determined); cf. [4].

It was shown in [24] that such an $x_\lambda(t)$ must be "very small,"

$$x_\lambda(t) = O(t^{-N}) \text{ for every fixed } N, \qquad (t \rightarrow \infty),$$

as soon as (35) is of *Grenzpunkt type by virtue of* (62). It turns out that, if (62) is replaced by another *Grenzpunkt* criterion, then not only can $x_\lambda(t)$ fail to be "very small," and not only can (39) become false, but even the following situation is possible: (35) is of *Grenzpunkt* type and has no essential spectrum, but the (L^2)-solution $x_\lambda(t)$, which exists for $-\infty < \lambda < \infty$, fails to be $O(1)$ at *any* λ.

This is accomplished by the proof of (c) in (v). In fact, the $f(t)$ constructed there was such as to make (35) non-oscillatory for every λ. Since this means that $\lambda^* = \infty$, it implies that (35) is of *Grenzpunkt* type and has no essential spectrum; cf. [2]. Finally, it was seen in the proof of (c) in (v) that (39) is violated at every λ.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] M. Bôcher, "On regular singular points of linear differential equations of the second order whose coefficients are not necessarily analytic," *Transactions of the American Mathematical Society*, vol. 1 (1900), pp. 40-53.

[2] P. Hartman, "On differential equations with non-oscillatory eigenfunctions," *Duke Mathematical Journal*, vol. 15 (1948), pp. 697-709.

[3] ———, "On the spectra of slightly disturbed linear oscillators," *American Journal of Mathematics*, vol. 71 (1949), pp. 71-79.

[4] ——— and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[5] ——— and A. Wintner, "On non-conservative linear oscillators of low frequency," *American Journal of Mathematics*, vol. 70 (1948), pp. 529-539.

[6] ——— and A. Wintner, "A criterion for the non-degeneracy of the wave equation," *ibid.*, vol. 71 (1949), pp. 206-213.

[7] ——— and A. Wintner, "On the location of spectra of wave equations," *ibid.*, vol. 71 (1949), pp. 214-217.

[8] ——— and A. Wintner, "On the Fourier-Laplace transcendents," *ibid.*, vol. 71 (1949), pp. 367-372.

[9] O. Haupt, "Ueber lineare homogene Differentialgleichungen 2. Ordnung mit periodischen Koeffizienten," *Mathematische Annalen*, vol. 79 (1919), pp. 278-285.

[10] A. Kneser, "Untersuchungen über die reellen Nullstellen der Integrale linearer Differentialgleichungenssysteme," *ibid.*, vol. 32 (1893), pp. 409-435.

[11] ———, "Untersuchung und asymptotische Darstellung der Integrale gewisser linearer Differentialgleichungen bei grossen reellen Werthen des Arguments, I," *Journal für die reine und angewandte Mathematik*, vol. 116 (1896), pp. 178-212.

[12] H. A. Kramers, "Das Eigenwertproblem im eindimensional periodischen Kraftfelde," *Physica*, vol. 2 (1935), pp. 483-490.

[13] H. Milloux, "Sur l'équation differentielle $x'' + xA(t) = 0$," *Prace Matematyczno-Fizyczne*, vol. 41 (1934), pp. 39-54.

[14] O. Perron, "Ueber ein vermeintliches Stabilitätskriterium," *Göttinger Nachrichten*, 1930, pp. 28-29.

[15] E. Schrödinger, "Verwaschene Eigenwertsspectra," *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, 1929, pp. 668-682.

[16] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit singulären Stellen und ihre Eigenfunktionen, *Göttinger Nachrichten*, 1909, pp. 37-63.

[17] ———, "Ueber gewönliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1909), pp. 220-269.

[18] A. Wintner, "The adiabatic linear oscillator," *American Journal of Mathematics*, vol. 68 (1946), pp. 385-397.

[19] ———, "(L^2)-connections between the potential and kinetic energies of linear systems," *ibid.*, vol. 69 (1947), pp. 5-13.

[20] ———, "Asymptotic integrations of the adiabatic oscillator," *ibid.*, vol. 69 (1947), pp. 251-272.

[21] ———, "On the normalization of the characteristic differentials in continuous spectra," *Physical Review*, vol. 72 (1947), pp. 516-517.

[22] ———, "Stability and high frequency," *Journal of Applied Physics*, vol. 18 (1947), pp. 941-942.

[23] ———, "A criterion of oscillatory stability," *Quarterly of Applied Mathematics*, vol. 7 (1949), pp. 115-117.

[24] ———, "On the smallness of isolated eigenfunctions," *American Journal of Mathematics*, vol. 71 (1949), pp. 603-611.

A SEPARATION THEOREM FOR CONTINUOUS SPECTRA.*

By PHILIP HARTMAN and AUREL WINTNER.

On the half-line $0 \leqq t < \infty$, let $p = p(t) > 0$ and $q = q(t)$ be real-valued, continuous functions, and consider real-valued, non-trivial ($\not\equiv 0$) solutions of the differential equation

$$(1) \qquad (px')' + (q + \lambda)x = 0,$$

where λ is a real parameter. Suppose that (1) is of *Grenzpunkt* type, i. e., that (1) does not possess two linearly independent solutions both of which are of class (L^2), that is to say that

$$(2) \qquad \int_0^\infty x^2(t)\,dt = 1.$$

According to Weyl [4], p. 238, this requirement will be satisfied for every λ if it is satisfied for a single λ. Since what is required is that not every solution of (1) be of class (L^2), the assumption requires that (1) and a linear, homogeneous boundary condition at $t = 0$, such as $x(0) = 0$ or $x'(0) = 0$ (more generally,

$$(3\alpha) \qquad x(0)\sin\alpha - p(0)x'(0)\cos\alpha = 0,$$

where α is a real number) should *determine* an eigenvalue problem.

In particular, there belongs to every choice of α in (3) a spectrum, $S(\alpha)$, and in it a point spectrum, $P(\alpha)$ (which can be vacuous, whilst $S(\alpha)$ cannot). According to Weyl [4], p. 251, the set of the cluster points of $S(\alpha)$ is independent of α and can therefore be denoted simply by S'. Thus $S(\alpha) = S' + P(\alpha)$, where it is understood that S' and $P(\alpha)$ can have points in common. Clearly, $P(\alpha) = P(\alpha + \pi)$; cf. (3). Conversely, a point λ belongs to both $P(\alpha)$ and $P(\beta)$ only if $\alpha \equiv \beta \pmod{\pi}$. For, on the one hand, a given λ is in $P(\alpha)$ if and only if (1) has, for this λ, an (L^2)-solution satisfying (3α) and, on the other hand, (1) should not have, for this (or for any) λ, two linearly independent (L^2)-solutions.

Let S^* denote the set of those points of the line $-\infty < \lambda < \infty$ which are not in S'. Then, since S' is a closed set (possibly the entire line), the

---

* Received October 25, 1948.

set S^* is open (possibly vacuous). It was shown in [2] that, if λ is in S^*, then (1) has a solution satisfying (2). This means that, if S^0 denotes the interior of (i. e., the greatest open set contained in) the set of those λ-values corresponding to which (1) has a solution satisfying (2), then S^* is a subset of S^0. We did not succeed in proving that S^0 is always identical with S^*. The nature of the difficulties involved will be clear from the type of the results which will be proved in the Appendix below.

Since S^* is in S^0, there belongs to every λ contained in S^* an $\alpha = \alpha_\lambda$ for which λ becomes a point of $P(\alpha)$, and this $\alpha = \alpha_\lambda$ is unique mod π (for every λ in S^*). Suppose that S^* is not vacuous (i. e., that not every real λ is in $S(\alpha)$ for some and/or every α), and let $S^* = \Sigma J_k$ be the canonical decomposition of the open set S^* into a (finite or infinite) sequence of mutually disjoint, open intervals (one of which can be a half-line). It will be shown below that, if the indeterminacy (mod π) of α is suitably normalized at every point λ of S^*, then

(i) *the* $\alpha = \alpha_\lambda$ *for which* λ *becomes a point of* $P(\alpha)$ *is a continuous function of the position,* λ*, on* $S^* = \Sigma J_k$*; what is more,*

(ii) *the continuous* α_λ *is a regular function on every* J_k*; in addition,*

(iii) *the continuous* α_λ *is a (strictly) increasing function on every* J_k.

At the end of the paper, a fact corresponding to (ii) will result concerning the eigenfunctions themselves. It is understood that, in (ii), the regularity of α_λ on the real λ-interval J_k means that α_λ possesses a (direct and regular) analytic continuation from J_k into the upper and lower half-planes of the complex variable λ.

Clearly, (i) and (iii) imply the following

SEPARATION THEOREM. *If* λ' *and* λ''*, where* $\lambda' < \lambda''$*, are (isolated) eigenvalues of any boundary condition* (3) *and if the open interval* (λ', λ'') *does not contain any point of the spectrum of that boundary condition, then* (λ', λ'') *contains exactly one point of the spectrum of every boundary condition* (3) *distinct from the given one.*

The need for such a separation theorem, a theorem depending on the monotony of α_λ, was formulated in [2] as a desideratum. The use of a classical argument of Sturm (cf. [4], pp. 252-255) is prevented by the possibility that two (λ', λ'')-intervals can be separated by points of S', e. g., by a continuous spectrum. That such a case of "band spectra" can actually occur, is shown by the example of any periodic $q(t) \not\equiv$ const. in (1), with $p(t) \equiv 1$.

Proof of (ii) and (iii).

(i) is contained in (ii). In order to prove (ii) and (iii), it will be sufficient to show that, if λ_0 is a given point of S^*, the function α_λ is regular and strictly increasing on *some*, sufficiently small, open interval containing λ_0.

Since (ii) requires the consideration of complex λ-values also, let (2) be reworded as

$$\int_0^\infty |x(t)|^2 \, dt = 1. \tag{4}$$

According to Weyl [4], p. 238, there belongs to every non-real λ a solution of (1) satisfying (4), and this solution is unique to a constant factor, (1) being of *Grenzpunkt* type.

Let λ_0 be a given point of S^*. Denote by α^0 an α-value corresponding to which $P(\alpha)$ contains λ_0, and normalize α^0 by $-\frac{1}{2}\pi \leqq \alpha^0 < \frac{1}{2}\pi$. Let θ be any real number satisfying $\theta \not\equiv \alpha^0 \pmod{\pi}$. Then, since λ_0 is in the (open) complement of S', a sufficiently small open interval containing λ_0 will not contain any point of $S(\theta)$. For, on the one hand, λ_0 is in $P(\alpha^0)$, where $\alpha \not\equiv \theta \pmod{\pi}$ and, on the other hand, $S(\theta)$, hence $P(\theta)$, does not cluster at λ_0. Hence, there exists about the point λ_0 of the complex λ-plane a sufficiently small circle having the property that, corresponding to every λ within the circle, (1) has a solution satisfying (4), and this solution is unique to a constant factor and does not satisfy (3θ).

It follows that, if λ is any point of a sufficiently small circle about λ_0, there exists a unique (complex) number $l = l(\lambda)$ having the property that

$$x = \xi(t, \lambda) \tag{5}$$

satisfies the differential equation (1), the (L^2)-condition

$$0 < \int_0^\infty |\xi(t, \lambda)|^2 \, dt < \infty \tag{6}$$

and the initial conditions

$$\xi(0, \lambda) = -\sin\theta + l(\lambda)\cos\theta, \qquad p(0)\xi'(0, \lambda) = \cos\theta + l(\lambda)\sin\theta, \tag{7}$$

where θ is the real constant fixed above. In fact, this definition of $l(\lambda)$ is equivalent to the following representation of the (L^2)-solution (5):

$$\xi(t, \lambda) = x_1(t, \lambda) + l(\lambda)x_2(t, \lambda), \tag{8}$$

where x_1, x_2 is that pair of linearly independent solutions of (1) which is determined by the initial conditions

$$(9_1) \qquad x_1(0,\lambda) = -\sin\theta, \qquad p(0)x'_1(0,\lambda) = \cos\theta;$$

$$(9_2) \qquad x_2(0,\lambda) = \cos\theta, \qquad p(0)x'_2(0,\lambda) = \sin\theta.$$

According to Weyl [4], pp. 226-227, the function $l(\lambda)$ or the resulting analytic continuation is regular in the *open* upper (and lower) λ-half-plane. Clearly, $l(\lambda)$ is real-valued along the real diameter of the circle.

In contrast to the above definition of (5), let

$$(10) \qquad x = x(t,\lambda)$$

denote that solution of (1) which belongs to the initial conditions

$$(11) \quad x(0,\lambda) = -\sin\theta + l(\lambda_0)\cos\theta, \qquad p(0)x'(0,\lambda) = \cos\theta + l(\lambda_0)\sin\theta.$$

Thus $x(t,\lambda_0) = \xi(t,\lambda_0)$, by (7), but, in contrast to the function (5), the function (10) is not in general of class $L^2(0,\infty)$. It becomes of class $L^2(0,\infty)$ at $\lambda = \lambda_0, \lambda_1, \lambda_2, \cdots$, where the latter sequence (which can be finite but, since it contains λ_0, cannot be vacuous) consists of the points of the point spectrum $P(\alpha^0)$. Put

$$(12) \qquad \psi_n{}^2 = \int_0^\infty x^2(t,\lambda_n)\,dt, \text{ where } 0 < \psi_n < \infty \qquad (n \geqq 0).$$

Besides the eigenfunctions $x(t,\lambda_n)$, use will have to be made of Hellinger's eigendifferentials $x(t,\lambda)\,d\phi(\lambda)$ ([3], pp. 252-258) which, according to Weyl ([4], pp. 239-251), can be introduced, with reference to (1), as follows: There exists on the line $-\infty < \lambda < \infty$ a continuous, non-decreasing function $\phi = \phi(\lambda)$ which is constant on each of the components of the open set S^* and has the following properties: For every fixed μ, where $-\infty < \mu < \infty$, the function of t

$$(13) \qquad \Phi(t,\mu) = \int_0^\mu x(t,\lambda)\,d\phi(\lambda)$$

is of class $L^2(0,\infty)$, satisfies the homogeneous integro-differential equation

$$(14) \qquad \{p(t)\Phi'(t,\mu)\}' + \{q(t)+\mu\}\Phi(t,\mu) = \int_0^\mu \Phi(t,\lambda)\,d\phi(\lambda)$$

[which corresponds to (1) if, besides $\lambda=\mu$, an infinitesimal vicinity of this $\lambda=\mu$ is also taken into account] and, when combined with the eigenfunctions, supplies a complete ortho-normal set. According to Hellinger (loc. cit.), the last condition means that, if $f(t)$ is any continuous function (or Baire function) of class $L^2(0,\infty)$, then Parseval's relation,

$$(15)\qquad \int_0^\infty |f(t)|^2\,dt=\sum_n |c(\lambda_n)|^2+\int_{-\infty}^{\infty} |dC(\lambda)|^2/d\phi(\lambda),$$

holds for the Fourier constants belonging to the contributions of the point spectrum and the continuous spectrum together, i. e., for

$$(16)\qquad c(\lambda_n)=\int_0^\infty f(t)x(t,\lambda_n)\,dt/\psi_n \qquad \text{[cf. (2), (12)]}$$

and

$$(17)\qquad C(\lambda)=\int_0^\infty f(t)\Phi(t,\lambda)\,dt.$$

Parseval's relation, (15), will be needed for the $f(t)$ represented by (5) (for a fixed $\lambda\neq\lambda_0$). In order to calculate the corresponding Fourier constants, (16) and (17), use will be made of the following fact: If D denotes the differential operator of the case $\lambda=0$ of (1), i. e., if

$$(18)\qquad D(f)=(pf')'+qf,$$

and if f, g are of class (L^2) on $0\leqq t<\infty$, have continuous derivatives f', g' which make the products pf', pg' absolutely continuous locally and the functions $D(f)$, $D(g)$ of class (L^2) on $0\leqq t<\infty$, then

$$(19)\qquad w_{fg}(t)\rightarrow 0 \text{ as } t\rightarrow\infty,$$

where $w_{fg}(t)$ denotes the Wronskian,

$$(20)\qquad w_{fg}=(f'g-g'f)p.$$

First, if $x=f$ and $x=g$ are real-valued and, besides satisfying the conditions just mentioned, satisfy *a common boundary condition*, (3α), then (19) follows from the relation

$$(21)\qquad \int_0^\infty fD(g)\,dt=\int_0^\infty gD(f)\,dt$$

and from the usual Green identity. The relation (21) was proved by Weyl [4], p. 242; it expresses the self-adjoint character of the problem. Next, if the last italicized proviso, without which (21) is not valid, is omitted, then (19) remains true, since, in view of (20), the relation (19) has nothing to do with the behavior of $f(t)$, $g(t)$ near $t=0$. Finally, it is clear from (20) that, (19) being true for real-valued f, g, it must be true for complex-valued f, g also.

Let (19) be applied to the pair of functions

$$f(t)=\xi(t,\lambda), \qquad g(t)=\Phi(t,\mu) \tag{22}$$

(which, in view of the defining properties of the functions (5), (13), is permissible). Since (5) and (13) are solutions of (1) and (14) respectively, (22), (20) and an application of Green's identity give

$$(\mu-\lambda)\int_0^\infty \xi(t,\lambda)\Phi(t,\mu)\,dt=0-w_{fg}(0)+\int_0^\infty \xi(t,\lambda)\Big(\int_0^\mu \Phi(t,\nu)\,d\nu\Big)dt,$$

the first term, 0, on the right being $w_{fg}(\infty)$, by (19). In order to evaluate the second term, $-w_{fg}(0)$, on the right, it is sufficient to observe that, on the one hand, the function $\Phi(t,\mu)$ of t satisfies the same homogeneous boundary condition as does the function (10), and that, on the other hand, the first of the functions (22) satisfies the initial condition (7). Hence it is clear from (11) that (20) reduces at $t=0$ to

$$w_{fg}(0)=\{l(\lambda_0)-l(\lambda)\}\phi(\mu)$$

in the case (22).

In view of the last two formula lines, the "continuous" Fourier constant, (17), of $f(t)=\xi(t,\lambda)$ is subject to the identity

$$(\mu-\lambda)C(\mu)=\{l(\lambda)-l(\lambda_0)\}\phi(\mu)+\int_0^\mu C(\nu)\,d\nu.$$

But $\Phi(t,0)\equiv 0$, by (13), hence $C(0)=0$, by (17), and so a partial integration gives

$$(\mu-\lambda)C(\mu)-\int_0^\mu C(\nu)\,d\nu=\int_0^\mu(\nu-\mu)\,dC(\nu).$$

Consequently,

$$\int_0^\mu(\nu-\lambda)\,dC(\nu)=\{l(\lambda)-l(\lambda_0)\}\phi(\mu).$$

This relation means that the Stieltjes-Hellinger differentials of the "continuous Fourier constants," (17), of $f(t)=\xi(t,\lambda)$ are given by

$$(\mu-\lambda)\,dC(\mu)=\{l(\lambda)-l(\lambda_0)\}\,d\phi(\mu).$$

A similar calculation shows that the corresponding "discrete" Fourier constants, (16), of $f(t)=\xi(t,\lambda)$ are given by

$$(\lambda_n-\lambda)\,c(\lambda_n)=\{l(\lambda)-l(\lambda_0)\}/\psi_n,$$

the normalizing factor, $1/\psi_n$, on the right being supplied by the integral (12). Accordingly, (15) becomes

$$\int_0^\infty|\xi(t,\lambda)|^2dt=|l(\lambda)-l(\lambda_0)|^2\{\sum_n|\lambda_n-\lambda|^{-2}/\psi_n^2+\int_{-\infty}^\infty|\mu-\lambda|^{-2}d\phi(\mu)\}.$$

In addition, it is clear from the above evaluations of the terms of Parseval's relation of $f(t)=\xi(t,\lambda)$ that

$$\int_0^\infty|x(t,\lambda_0)+(\lambda-\lambda_0)\psi_0\xi(t,\lambda)/\{l(\lambda)-l(\lambda_0)\}|^2\,dt$$

$$=|\lambda-\lambda_0|^2\psi_0^2\{\sum_{\lambda_n\neq\lambda_0}|\lambda_n-\lambda|^{-2}/\psi_n^2+\int_{-\infty}^\infty|\mu-\lambda|^{-2}d\phi(\mu)\}.$$

The last relation supplies the basis for the proof of (ii) and (iii).

First, since the eigenvalues λ_n do not cluster at λ_0, and since $\phi(\lambda)$ is constant on an open interval containing λ_0, the expression on the right of the last equality tends to 0 as $\lambda\to\lambda_0$. Hence, the same is true of the integral on the right. Consequently, as $\lambda\to\lambda_0$,

$$(23)\qquad \int_0^T|x(t,\lambda_0)+(\lambda-\lambda_0)\psi_0\xi(t,\lambda)/\{l(\lambda)-l(\lambda_0)\}|^2\,dt\to0$$

holds for every positive T. From this, it will now be deduced that

$$(24)\qquad l^*(\lambda_0)=-1/\psi_0,\text{ where }l^*(\lambda)=dl(\lambda)/d\lambda$$

(whilst $'=d/dt$).

Proof of (24). Two cases will be distinguished, according as there does or does not exist a γ corresponding to which it is possible to choose a sequence of real or complex numbers β_k satisfying

$$l(\beta_k) - l(\lambda_0) \to \gamma \text{ and } \beta_k \to \lambda_0, \text{ as } k \to \infty. \tag{25}$$

In both cases, use will be made of the identity

$$\xi(t, \lambda) - x(t, \lambda) = \{l(\lambda) - l(\lambda_0)\} x_2(t, \lambda) \tag{26}$$

which, since (5) and (10) are solutions of (1), and since a solution $x(t)$ of (1) is uniquely determined by a pair of initial constants $x(0)$, $x'(0)$, is clear from (7), (9_2) and (11).

In the first case, that is, if there exist values γ, β_k satisfying (25), the identity (26), when combined with the local continuity of the general solution of (1), shows that, as $k \to \infty$, the limit relation

$$\xi(t, \beta_k) \to x(t, \lambda_0) + \gamma x_2(t, \lambda_0)$$

holds uniformly on every fixed interval $0 \leqq t \leqq T < \infty$. In view of (23), this implies that, for every positive $T < \infty$,

$$\lim_{k\to\infty} \int_0^T |\xi(t, \beta_k)|^2 \, dt \text{ exists } (\neq \infty).$$

It follows therefore from (25) and from the case $\lambda = \beta_k$ of (23), where $\lambda \to \lambda_0$, that either the value of the constant γ is 0 or

$$\int_0^T |x(t, \lambda_0)|^2 \, dt = 0$$

holds for every T. In other words, either $\gamma = 0$ or $x(t, \lambda_0) \equiv 0$. Since $x(t, \lambda_0)$ is an eigenfunction, it follows that $\gamma = 0$.

In the second case, specified before (25), there does not exist any γ corresponding to which it is possible to choose a sequence $\beta_1, \beta_2, \cdots$ satisfying (25). Hence, the second case is characterized by $|l(\lambda) - l(\lambda_0)| \to \infty$ as $\lambda \to \lambda_0$, where λ varies freely (near λ_0). It follows therefore from (26) that $x_2(t, \lambda) \to 0$, as $\lambda \to \lambda_0$, holds for every t. But this means that $x_2(t, \lambda_0) \equiv 0$ and contains, therefore, a contradiction.

Accordingly, the second case is impossible, and the first case leads to $\gamma = 0$. In view of the definition of the two cases, specified after (25), it now follows that, when λ varies freely, $l(\lambda) - l(\lambda_0) \to 0$ as $\lambda \to \lambda_0$. Consequently, from (26),

$$\int_0^T |x(t, \lambda) - \xi(t, \lambda)|^2 \, dt \to 0$$

as $\lambda \rightarrow \lambda_0$. Hence it is clear from (2) that

$$(\lambda - \lambda_0)\psi_0/\{l(\lambda) - l(\lambda_0)\} \rightarrow -1 \text{ as } \lambda \rightarrow \lambda_0.$$

This proves (24), since $\psi_0 \neq 0$, by (12).

Completion of the proof of (ii) *and* (iii). Since the sequence $\beta_1, \beta_2, \cdots$ occurring in (25) was allowed to be complex, the above proof of (24) implies that the derivative of $l(\lambda)$ at the (real) point $\lambda = \lambda_0$ exists even if $\Delta\lambda$ in $[l(\lambda_0 + \Delta\lambda) - l(\lambda_0)]/\Delta\lambda$, where $\Delta\lambda \rightarrow 0$, is complex. In other words, the function $l(\lambda)$ of the complex variable λ is differentiable at the point $\lambda = \lambda_0$ of the real axis. On the other hand, $l(\lambda)$ is differentiable at every non-real λ, since it is clear from Weyl's theory ([4], pp. 226-227) that $l(\lambda)$ is regular in the open upper and lower halves of the λ-plane. But λ_0 was chosen to be any real number not contained in S'. Consequently, $l(\lambda)$ is differentiable at every point of a circle in the complex λ-plane having as its center the real point $\lambda = \lambda_0$. Hence, $l(\lambda)$ is regular at $\lambda = \lambda_0$.

From now on, λ will be assumed to be real and near λ_0. Then $l(\lambda)$ is real-valued. Furthermore, the function $l(\lambda)$ is decreasing at $\lambda = \lambda_0$, by (24) and (12). Since $l(\lambda)$ is regular at $\lambda = \lambda_0$, it is decreasing near $\lambda = \lambda_0$ also. Finally, $l(\lambda)$ is continuous at and near $\lambda = \lambda_0$. It follows that if α^0 and θ are the constants chosen, with reference to λ_0, at the beginning of the section which follows (4), then

$$(27) \qquad \alpha_\lambda = \tfrac{1}{2}\pi + \theta - \operatorname{arc\,tan} l(\lambda)$$

and the initial determination $\alpha_{\lambda_0} = \alpha^0$ define, on an open interval about λ_0, a unique continuous function, α_λ.

According to (27) and (7),

$$\xi(0, \lambda) \sin \alpha_\lambda - p(0)\xi'(0, \lambda) \cos \alpha_\lambda = 0.$$

This means that the solution (5) of (1) satisfies the case $\alpha = \alpha_\lambda$ of the boundary condition (3a). Furthermore, the solution (5) satisfies (6); so that, since (5) is now real-valued,

$$(12\text{ bis}) \qquad \psi^2(\lambda) = \int_0^\infty \xi^2(t, \lambda)\, dt, \text{ where } 0 < \psi(\lambda) < \infty,$$

defines near λ_0 a function, $\psi(\lambda)$, satisfying

$$(28) \qquad d\alpha_\lambda/d\lambda = \{1 + l^2(\lambda)\}^{-1}/\psi(\lambda) > 0.$$

In fact, (28) follows from (27) and (24), if the λ_0 of (24), which is any point of S^*, is called λ.

Since $l(\lambda)$ is regular at λ_0, it is seen from (28), and from the remark made before (4), that the proof of (ii) and (iii) is now complete.

Remark. The last two formula lines show that, if (5) is replaced by its constant multiple

$$x^\lambda(t) = \{[1 + l^2(\lambda)]d\alpha_\lambda/d\lambda\}\xi(t,\lambda),$$

then, besides (1) and ($3\alpha_\lambda$), (2) will be satisfied by $x(t) = x^\lambda(t)$. It is easily realized that not only the normalizing constant, $x^\lambda(t)/\xi(t,\lambda)$, is regular function on every λ-interval J_k but that the same holds for the (L^2)-solution $x^\lambda(t)$ itself, if the latter is thought of as a function of λ while t is fixed on the half-line $0 \leqq t < \infty$ of (1).

Appendix.

As above, let S^0 denote the set of those real numbers λ^0 corresponding to which the differential equation ($\overset{\circ}{1}$), which is supposed to be of *Grenzpunkt* type, has a solution satisfying (2) whenever λ is in a sufficiently small interval $(\lambda^0 - \epsilon, \lambda^0 + \epsilon)$.

With reference to a fixed choice of the boundary condition (3α), let $S(\alpha)$, $P(\alpha)$ and $Q(\alpha)$ respectively denote the spectrum, the point spectrum and the continuous spectrum (the latter is defined as the complement of the set of the greatest open λ-intervals on which, in the above notations, Hellinger's continuous function $\phi(\lambda) = \phi_\alpha(\lambda)$ is constant). Thus $S(\alpha) = Q(\alpha) + P(\alpha) + P'(\alpha)$ and $S'(\alpha) = Q(\alpha) + P'(\alpha)$, if A' denotes the set of the cluster points of a λ-set A.

In connection with his proof of the fact that $S'(\alpha)$ is independent of α, Weyl [4], pp. 251-252 points out that each of the two terms, $Q(\alpha)$ and $P'(\alpha)$, of $S' = S'(\alpha)$ might be, but has not been proved to be, independent of α. While this question will remain unanswered, a partial result can be deduced by the method used above. In fact, the following theorem will be proved:

(I) *The open set S^0 does not contain any point of any of the continuous spectra $Q(\alpha)$, $-\frac{1}{2}\pi \leqq \alpha < \frac{1}{2}\pi$, which are determined by a boundary condition* (3α) *and a differential equation* (1) *of Grenzpunkt type.*

In view of the parallel standing of $Q(\alpha)$ and $P'(\alpha)$ in S', it is natural to expect that (I) should remain true if $Q(\alpha)$ is replaced by $P'(\alpha)$. But all that the method leading to (I) will supply in this regard can be formulated as follows:

(II) *In the notations of* (I), *the closure of any of the point spectra* $P(\alpha)$, *where* $-\frac{1}{2}\pi \leqq \alpha < \frac{1}{2}\pi$, *is nowhere dense on the open set* S^0; *what is more, no subset of any of the intersections* $P(\alpha)S^0$ *is dense on itself.*

In order to prove (I) and (II), let an α^0, where $-\frac{1}{2}\pi \leqq \alpha^0 < \frac{1}{2}\pi$, be arbitrarily fixed, and let $x = x(t;\lambda)$, where $-\infty < \lambda < \infty$, denote that solution of (1) which is determined by the initial conditions

$$x(0;\lambda) = \cos\alpha^0, \qquad p(0)x'(0;\lambda) = \sin\alpha^0.$$

Clearly, the solution $x(t;\lambda)$ thus defined differs from the above solution $x(t,\lambda)$, defined by (11), only in a constant factor. In particular, $x = x(t;\lambda)$ satisfies ($3\alpha^0$). Hence, if $\lambda_0, \lambda_1, \cdots$ denotes the (finite or infinite) sequence which forms the point spectrum $P(\alpha^0)$, then the respective eigenfunctions can be chosen to be $x(t;\lambda_0), x(t;\lambda_1), \cdots$. Similarly, the eigendifferentials corresponding to the points λ of $Q(\alpha^0)$ are of the form $x(t;\lambda)d\phi(\lambda)$, where the monotone function $\phi(\lambda)$ depends on the choice of α^0. Obviously, formulae (12)-(17) remain applicable.

Let λ be a point of the open set S^0, defined above. Then (1) has a real-valued, non-trivial (L^2)-solution. Let this solution, which is unique to a constant factor, be denoted by $x = \xi(t;\lambda)$. Since the values of ξ and ξ' at $t = 0$ can be multiplied by any real, non-vanishing constant, they can be normalized by

$$\xi(0;\lambda) = \cos\alpha, \qquad p(0)\xi'(0;\lambda) = \cos\alpha,$$

where $\alpha = \alpha(\lambda)$ is unique mod π. Clearly, the solution $x = \xi(t;\lambda)$ thus defined differs from the above solution $x = \xi(t,\lambda)$, defined by (7), only in a constant factor.

Let Parseval's relation, (15), be applied to $f(t) = \xi(t;\lambda)$ (instead of to $f(t) = \xi(t,\lambda)$, as above). To this end, it is sufficient to observe that, in view of the last two formula lines, the value of the Wronskian (20) at $t = 0$ now reduces to

$$w_{fg}(0) = \phi(\mu)\sin(\alpha - \alpha^0), \text{ with } \alpha = \alpha(\lambda),$$

if (22) is replaced by $f(t) = \xi(t;\lambda)$, $g(t) = \Phi(t,\mu)$; cf. (13), where $x(t,\lambda)$ must be replaced by $x(t;\lambda)$. Correspondingly, it is seen from the above calculation of the Fourier constants, (16) and (17), of $f(t) = \xi(t,\lambda)$ that the Fourier constants of $f(t) = \xi(t;\lambda)$ are given by

$(\lambda_n - \lambda) c(\lambda_n) = \{\sin(\alpha^0 - \alpha)\}/\psi_n, \quad (\mu - \lambda) dC(\mu) = \{\sin(\alpha^0 - \alpha)\} d\Phi(\mu).$

Consequently, the explicit form of the case $f(t) = \xi(t;\lambda)$ of (15) can be written as

$$\int_0^\infty \xi^2(t;\lambda)\,dt/\sin^2(\alpha^0 - \alpha) = A(\lambda) + B(\lambda)$$

if $\sin(\alpha^0 - \alpha) \neq 0$, where λ is any point of S^0, and $A(\lambda)$, $B(\lambda)$ are defined by

$$A(\lambda) = \int_{-\infty}^{\infty} (\mu - \lambda)^{-2} d\phi(\mu), \qquad B(\lambda) = \sum_n (\lambda_n - \lambda)^{-2}/\psi_n^2.$$

It follows that $A(\lambda) < \infty$ and $B(\lambda) < \infty$, provided that the point λ of S^0 is such that the corresponding $\alpha = \alpha(\lambda)$, defined above, satisfies the proviso $\sin(\alpha^0 - \alpha) \neq 0$. Clearly, this proviso, $\alpha(\lambda) \not\equiv \alpha^0 \pmod{\pi}$, is satisfied if and only if λ is not in $P(\alpha^0)$. Accordingly, $A(\lambda) < \infty$ and $B(\lambda) < \infty$ hold for all those points λ of S^0 which are not in $P(\alpha^0)$. From this result, it will now be easy to deduce both (I) and (II).

Proof of (I). Let $S^0 = \sum J^i$ be the decomposition of the open set S^0 into mutually disjoint, open intervals J^i (one of which can be a half-line), and let $J = J(\lambda)$ denote that J^i which contains a given point, λ, of S^0. The assertion of (I) is equivalent to the statement that the monotone function $\phi(\mu)$, occurring in the above definition of $A(\lambda)$, is constant on the μ-interval J. In fact, while it is true that the function $\phi(\mu)$ depends on the choice of the $\alpha = \alpha^0$ which determines the continuous spectrum $Q(\alpha) = Q(\alpha^0)$, the angle α^0 has above been chosen arbitrarily.

Suppose that the point λ of J is not in $P(\alpha^0)$. Then $A(\lambda) < \infty$. Hence, by the definition of $A(\lambda)$, and since $\phi(\mu)$ is monotone,

$$\infty > \left| \int_{\lambda}^{\lambda \pm \epsilon} (\mu - \lambda)^{-2} d\phi(\mu) \right| \geqq |\phi(\lambda \pm \epsilon) - \phi(\lambda)|/\epsilon^2,$$

where $\epsilon > 0$ is arbitrary. Consequently, if λ is not in $P(\alpha^0)$, the derivative $d\phi(\lambda)/d\lambda$ exists and is 0. Since $\phi(\mu)$ is monotone and continuous, and since the point spectrum $P(\alpha^0)$ is enumerable (at most), it follows, first, that $\phi(\lambda)$ is absolutely continuous and then, that $\phi(\lambda)$ is constant on J.

Proof of (II). If λ is in S^0 but not in $P(\alpha^0)$, then $B(\lambda) < \infty$. But $B(\lambda)$ was defined as a Borel series of the form $B(\lambda) = \sum c_n/(\lambda - \lambda_n)^2$, where the coefficients $c_n = 1/\psi_n^2$ are positive. Since the values λ_n occurring in the denominators form the sequence which constitutes the point spectrum $P(\alpha^0)$, it follows that the series $\Sigma c_n/(\lambda - \lambda_n)^2$ is convergent at all those points of the open set S^0 which are distinct from every λ_n. Hence, (II) follows from classical properties of the λ-set on which a Borel series must diverge; cf. e. g., [1], pp. 314-315.

Remark. Let $M = M(\alpha^0)$ denote the sequence, say $\mu_1 = \mu_1(\alpha^0)$, $\mu_2 = \mu_2(\alpha^0), \cdots$, of those eigenvalues $\lambda_n = \lambda_n(\alpha^0)$ which are contained in an open interval, J. If it could be proved that the corresponding Borel series, say $\Sigma b_n/(\lambda - \mu_n)^2$, where $b_n > 0$, cannot converge at every point of $J - M$ unless the set of the cluster points of M is enumerable (at most), it would follow that the set which (II) claims to be nowhere dense is enumerable as well. In view of (I), this would prove that the intersection of S^0 and $S' = Q(\alpha) + P'(\alpha)$, besides being nowhere dense, is enumerable (at most). It would still not follow that S^0 is entirely free of points of S'.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] H. Hahn, *Theorie der reellen Funktionen*, vol. I, Berlin, 1921.

[2] P. Hartman and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[3] E. Hellinger, "Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen," *Journal für Mathematik*, vol. 136 (1909), pp. 210-271.

[4] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

ADDITIVE GROUPS AND LINEAR MANIFOLDS OF TRANSFORMATIONS BETWEEN BANACH SPACES.[*] [1]

By BERTRAM YOOD.

1. Introduction. In this paper we consider the space $A(\mathfrak{X}, \mathfrak{Y})$ of all bounded linear transformations with domain $\mathfrak{X}$ and range in $\mathfrak{Y}$, where $\mathfrak{X}$ and $\mathfrak{Y}$ are Banach spaces with complex scalars. We study $A(\mathfrak{X}, \mathfrak{Y})$ in the uniform, strong, weak, finite and weakly finite topologies. It is our main purpose to obtain formulas for the closure of certain classes of additive groups and linear manifolds in $A(\mathfrak{X}, \mathfrak{Y})$ in these topologies (except the uniform). The first step in this direction is the demonstration in **2** that the closure of a set B in $A(\mathfrak{X}, \mathfrak{Y})$ is contained in a set related to B (which may vary with the topologies). Then it is shown that, under appropriate conditions, the closure of additive groups and linear manifolds attains this maximum possible size. To a large extent the arguments are algebraic. These are based on the theory of rings as developed by Jacobson in [5, 6, 7]. If the notion of ideal is defined in a natural way (5.1), the closure in the strong and weak topologies of each of these groups and linear manifolds is the intersection of the smallest weakly closed left and right ideals containing the given set.

In the use of Jacobson's ideas in this connection the following difficulty arises. Let B be a one-fold transitive ring in $A(\mathfrak{X}, \mathfrak{X})$. (We are using the terminology of Jacobson [5]). By the results of Jacobson [5], B is *dense*, in the sense that for any two sets $x_1, \cdots, x_n$; $y_1, \cdots, y_n$ of vectors where the members of the first set are linearly independent (with respect to the complex scalars) there exists a $T \varepsilon B$ such that $T(x_i) = y_i$, $i = 1, \cdots, n$, if and only if the centralizer of B in the set of all additive transformations defined on $\mathfrak{X}$ is the set of all complex scalar multiples of the identity transformation. We have been unable to verify or refute the statement that a one-fold transitive ring B here has both these properties. But we do show that this is true if B is an algebra closed in the uniform topology. This is sufficient to show that any one-fold transitive algebra in $A(\mathfrak{X}, \mathfrak{X})$ is dense in the strong and weak topologies of $A(\mathfrak{X}, \mathfrak{X})$.

---

[*] Received July 9, 1948.

[1] Presented to the American Mathematical Society, February 22, 1947, under the title *On ideals in operator rings over Banach spaces.* This is a revised version of a portion of the author's dissertation, Yale University, June 1947.

In **5** we consider ideals in $A(\mathfrak{X}, \mathfrak{Y})$ and certain quotient (difference) spaces which can be defined in terms of subgroups. In **6** we discuss the notion of centralizer.

The author wishes to express his gratitude to Professors Nelson Dunford and C. E. Rickart for their suggestions and encouragement in the writing of this paper.

2. Preliminaries.

Let $\mathfrak{X}$ and $\mathfrak{Y}$ be two Banach spaces and let $A(\mathfrak{X}, \mathfrak{Y})$ be the class of all bounded linear transformations with domain $\mathfrak{X}$ and range in $\mathfrak{Y}$. Under the usual definitions of addition and multiplication by complex scalars, $A(\mathfrak{X}, \mathfrak{Y})$ is a linear space. When $\mathfrak{X} = \mathfrak{Y}$ we have the additional operation of multiplication, $V = TU$ defined by the relation $V(x) = T[U(x)]$.

Following the ideas of von Neumann [10] we consider the uniform, strong and weak topologies for $A(\mathfrak{X}, \mathfrak{Y})$. The uniform topology is the Banach space topology obtained by defining a norm for transformations, $\| T \| = \sup \| T(x) \|$, $\| x \| \leqq 1$. A strong neighborhood of T_0 is any set of the form $N[T_0; x_1, \cdots, x_n; \epsilon]$ which is composed of all transformations T in $A(\mathfrak{X}, \mathfrak{Y})$ such that $\| T(x_k) - T_0(x_k) \| < \epsilon$, $k = 1, \cdots, n$ where $x_1, \cdots, x_n$ are elements of $\mathfrak{X}$ and $\epsilon > 0$. A weak neighborhood is any set $N[T_0; x_1, \cdots, x_n; y_1^*, \cdots, y_n^*; \epsilon]$ composed of transformations $T \varepsilon A(\mathfrak{X}, \mathfrak{Y})$ satisfying the relations $| y_k^*[T(x_k)] - y_k^*[T_0(x_k)] | < \epsilon$, $k = 1, \cdots, n$, where the x's and ϵ are described above and $y_1^*, \cdots, y_n^*$ are n arbitrary elements of $\mathfrak{Y}^*$, the conjugate space of $\mathfrak{Y}$. The finite topology is described by Jacobson [7, p. 11]. A finite neighborhood of T_0 is any set $N[T_0; x_1, \cdots, x_n]$ which consists of all $T \varepsilon A(\mathfrak{X}, \mathfrak{Y})$ such that $T(x_k) = T_0(x_k)$, $k = 1, \cdots, n$. A weakly finite neighborhood of T_0 is any set $N[T_0; x_1, \cdots, x_n; y_1^*, \cdots, y_n^*]$ of all $T \varepsilon A(\mathfrak{X}, \mathfrak{Y})$ such that $y_k^*[T(x_k)] = y_k^*[T_0(x_k)]$, $k = 1, \cdots, n$. All neighborhoods are obtained by varying T_0, ϵ, x_k, y_k^*, n over their domains of definition.

Let B be any subset of $A(\mathfrak{X}, \mathfrak{Y})$. Connected with B are two sets $\mathfrak{M}(B) \subset \mathfrak{X}$ and $\mathfrak{R}(B) \subset \mathfrak{Y}$ defined by

$$\mathfrak{M}(B) = \{x \varepsilon \mathfrak{X} \mid T(x) = 0 \text{ for every } T \varepsilon B\}$$

$$\mathfrak{R}(B) = \{y \varepsilon \mathfrak{Y} \mid y = T(x) \text{ for some } T \varepsilon B,\ x \varepsilon \mathfrak{X}\}.$$

We use the notation $\{x \mid P\}$ to mean the set of all x with the property P. $\mathfrak{M}(B)$ is necessarily a closed linear manifold of X. $\mathfrak{R}(B)$ has the property that if $y \varepsilon \mathfrak{R}(B)$ and λ is a scalar, then $\lambda y \varepsilon \mathfrak{R}(B)$. However examples can be readily supplied to show that $\mathfrak{R}(B)$ need not be a linear manifold, even if B is a linear manifold in $A(\mathfrak{X}, \mathfrak{Y})$. Consider, for instance, the linear manifold generated by two $1 - 1$ transformations whose ranges have only the zero element of Y in common. We let $\mathfrak{N}_1(B)$ designate the linear manifold in Y generated by $\mathfrak{N}(B)$.

2.1. THEOREM. *The closure of a set $B \subset A(\mathfrak{X}, \mathfrak{Y})$ in the finite, strong and weak topologies is contained in the following sets, respectively,*

$$\{T \varepsilon A(\mathfrak{X}, \mathfrak{Y}) \mid \mathfrak{M}(T) \supset \mathfrak{M}(B), \mathfrak{N}(T) \subset \mathfrak{N}(B)\} \tag{1}$$

$$\{T \varepsilon A(\mathfrak{X}, \mathfrak{Y}) \mid \mathfrak{M}(T) \supset \mathfrak{M}(B), \mathfrak{N}(T) \subset \overline{\mathfrak{N}(B)}\} \tag{2}$$

$$\{T \varepsilon A(\mathfrak{X}, \mathfrak{Y}) \mid \mathfrak{M}(T) \supset \mathfrak{M}(B), \mathfrak{N}(T) \subset \overline{\mathfrak{N}_1(B)}\}. \tag{3}$$

The closure of B in the weakly finite topology is contained in (3).

That the closure of B in the finite topology is contained in (1) follows directly from the definitions. Of all the topologies considered here, the weak topology is the weakest in the sense that the closure of a set in it is at least as large as is the closure in any of the other topologies. Suppose that $T_0(x) = y \neq 0$ for some $x \varepsilon \mathfrak{M}(B)$. Let $y^* \varepsilon \mathfrak{Y}^*$ have the property that $y^*(y) = 1$. Then no $T \varepsilon B$ is in the weak neighborhood $N[T_0; x; y^*; \frac{1}{2}]$ of T_0. Thus for T in the closure of B in any of the topologies, $\mathfrak{M}(T) \supset \mathfrak{M}(B)$.

If $T_0(x) = y$, $y \not\varepsilon \overline{\mathfrak{N}(B)}$ for some $x \varepsilon \mathfrak{X}$, then no $T \varepsilon B$ is in the strong neighborhood $N[T_0; x; \rho]$ of T_0, where ρ is the distance from y to $\mathfrak{N}(B)$. This proves the statement on the strong topology. Suppose that for some $x \varepsilon \mathfrak{X}$, $T_0(x) = y_0$, $y_0 \not\varepsilon \overline{\mathfrak{N}_1(B)}$. Then there exists a linear functional $y^* \varepsilon \mathfrak{Y}^*$ such that $y^*(y_0) = 1$ while $y^*(y) = 0$ for $y \varepsilon \overline{\mathfrak{N}_1(B)}$. No $T \varepsilon B$ is in the weak neighborhood $N[T_0; x; y^*; \frac{1}{2}]$ of T_0. This proves the statements for the weak and weakly finite topologies.

Next let G be an additive group of transformations in $A(\mathfrak{X}, \mathfrak{Y})$. Connected with G are a number of rings of transformations. Let

$$R_1(G) = \{T \varepsilon A(\mathfrak{Y}, \mathfrak{Y}) \mid TU \varepsilon G \text{ for every } U \varepsilon G\}$$

$$R_2(G) = \{T \varepsilon A(\mathfrak{X}, \mathfrak{X}) \mid UT \varepsilon G \text{ for every } U \varepsilon G\}.$$

It is readily verified that the group property of G implies that $R_1(G)$ and $R_2(G)$ are rings of transformations in $A(\mathfrak{Y}, \mathfrak{Y})$ and $A(\mathfrak{X}, \mathfrak{X})$ respectively.

Every $T \varepsilon R_1(G)$ has the property that $T(y) \varepsilon \mathfrak{N}(G)$ if $y \varepsilon \mathfrak{N}(G)$. For if otherwise $T(y_1) = y_2$, $y_1 \varepsilon \mathfrak{N}(G)$, $y_2 \not\varepsilon \mathfrak{N}(G)$, then by the definition of $\mathfrak{N}(G)$ there exists $U \varepsilon G$ and $x \varepsilon \mathfrak{X}$ such that $U(x) = y_1$, $TU(x) = y_2$ so that $TU \varepsilon G$, contrary to the definition of $R_1(G)$. Thus the image of $\mathfrak{N}_1(G)$ is contained in $\mathfrak{N}_1(G)$ for every $T \varepsilon R_1(G)$. Let T' be the contraction of T

to the domain $\mathfrak{N}_1(G)$. Then the mapping $T \rightarrow T'$ defined for $T \varepsilon R_1(G)$ is a ring homomorphism under which the image of $R_1(G)$ is again a ring which we shall call $R_l(G)$. $R_l(G)$ is a ring of transformations from $\mathfrak{N}_1(G)$ to $\mathfrak{N}_1(G)$.

By precisely the same reasoning, every $T \varepsilon R_2(G)$ takes $\mathfrak{M}(G)$ into $\mathfrak{M}(G)$. We consider the difference (quotient) space of $\mathfrak{X}$ modulo $\mathfrak{M}(G)$, $\mathfrak{X}/\mathfrak{M}(G)$. It is well known, see [2, p. 232], that if we define the norm $\| W \|$ of a coset W as $\| W \| = \inf \| x \|$, $x \varepsilon W$, then $\mathfrak{X}/\mathfrak{M}(G)$ is a Banach space. We use the notation x' to designate the coset of $\mathfrak{X}/\mathfrak{M}(G)$ which contains x. Since each $T \varepsilon R_2(G)$ vanishes on $\mathfrak{M}(G)$, each $T \varepsilon R_2(G)$ defines a transformation T' from $\mathfrak{X}$ to $\mathfrak{X}/\mathfrak{M}(G)$ by the rule $T'(x) = [T(x)]'$. Since $\| T'(x) \| \leqq \| T(x) \| \leqq \| T \| \, \| x \|$, $\| T' \| \leqq \| T \|$, and T' is bounded. Also as $T'(x) = 0'$ for $x \varepsilon \mathfrak{M}(G)$, we can define $T''(x') = T'(x)$ as a transformation from $\mathfrak{X}/\mathfrak{M}(G)$ to itself. For each $x \varepsilon x'$, $\| T''(x') \| \leqq \| T' \| \, \| x \|$. Clearly $\| T'' \| \leqq \| T' \|$ so that T'' is bounded. The correspondence $T \rightarrow T''$ has the properties $(T + U)'' = T'' + U''$; $(TU)'' = T''U''$. The latter property holds since $T''U''(x') = T''[U(x)]' = [TU(x)]' = (TU)''(x')$. Thus the correspondence $T \rightarrow T''$ takes the ring $R_2(G)$ into a ring $R_r(G)$ of bounded linear transformations from $\mathfrak{X}/\mathfrak{M}(G)$ to itself. $R_r(G)$ is defined unambiguously by G.

As a bridge between the closure in the finite topology and that in the strong and weak topologies we shall use the following result.

2.2. THEOREM. *Let $B \subset A(\mathfrak{X}, \mathfrak{Y})$ have the property that its closure in the finite topology is the set* (1) *of Theorem* 2.1. *Then its closure in the strong topology is the set* (2) *of Theorem* 2.1. *If $\mathfrak{N}(B)$ is a linear manifold, this is also its closure in the weak topology.*

Let T_0 be in the set (2) of Theorem 2.1 and let $N[T_0; x_1, \cdots, x_n; \epsilon]$ be a strong neighborhood of T_0. By renumbering, if necessary, we may suppose that $x_1, \cdots, x_r$ is a linearly independent subset of $x_1, \cdots, x_n$ which generates the same linear manifold. Then we can write

$$x_i = \sum_{k=1}^{r} a_{ik} x_k \qquad i = r + 1, \cdots, n.$$

Since $\mathfrak{N}(T_0) \subset \overline{\mathfrak{N}(B)}$ we can select vectors $y_i \varepsilon \mathfrak{N}(B)$ such that $\| y_i - T_0(x_i) \| < \min \{\epsilon, \epsilon/rA\}$ for $i = 1, \cdots, r$ where $A = \max | a_{ik} |$ for $r + 1 \leqq i \leqq n$ and $1 \leqq k \leqq r$. Now the closure, $\bar{B}$, of B in the strong topology contains its closure in the finite topology. Then by hypothesis we can select $T \varepsilon \bar{B}$ such that $T(x_i) = y_i$, $i = 1, \cdots, r$. For these values of i, $\| T(x_i) - T_0(x_i) \| < \epsilon$, and a simple computation shows that this is also true for $i = r + 1, \cdots, n$. Consequently $T_0 \varepsilon \bar{\bar{B}} = \bar{B}$. Then by Theorem 2. 1, $\bar{B}$ is the set (2) of Theorem 2. 1. Again by Theorem 2. 1, if $\mathfrak{N}(B)$ is a linear manifold, the set (2) is the closure in the weak topology.

We conclude this section with a necessary condition on strong closure.

2. 3 LEMMA. *Let $x_1, \cdots, x_n$ be n linearly independent vectors in a Banach space $\mathfrak{X}$. Then there exists a number $\epsilon > 0$ such that if $\| y_i - x_i \| < \epsilon$, $y_i \varepsilon \mathfrak{X}$, $i = 1, \cdots, n$, the vectors y_i are linearly independent.*

Let I be the identity transformation from $\mathfrak{X}$ to itself. Let x_i^* be linear functionals defined on $\mathfrak{X}$ such that $x_i^*(x_j) = \delta_{ij}$, $i, j = 1, \cdots, n$. We form the transformation

$$T(x) = I(x) + \sum_{i=1}^{n} x_i^*(x)(y_i - x_i).$$

It is readily verified that T is a bounded linear transformation and that

$$\| I - T \| \leqq \sum_{i=1}^{n} \| x_i^* \| \, \| y_i - x_i \| .$$

Thus if we select $\epsilon = \frac{1}{2}nA$ where $A = \max \| x_i^* \|$, $\| I - T \| < \frac{1}{2}$. From this it follows (see, for example, Gelfand [4, p. 4]) that T is an isomorphism. But $T(x_i) = y_i$, $i = 1, \cdots, n$ and thus the y_i's are linearly independent.

2. 4. THEOREM. *Let B be any set in $A(\mathfrak{X}, \mathfrak{Y})$. In order that T_0 be in the closure of B in the strong topology it is necessary that, whenever $x_1, \cdots, x_n$ in $\mathfrak{X}$ and $T_0(x_1), \cdots, T_0(x_n)$ in $\mathfrak{Y}$, each linearly independent sets of vectors, there exists a transformation $T \varepsilon B$ such that the vectors $T(x_1), \cdots, T(x_n)$ are linearly independent.*

Consider the ϵ of Lemma 2. 3 for the vectors $T_0(x_1), \cdots, T_0(x_n)$ of $\mathfrak{Y}$. Every T in the strong neighborhood $N[T_0; x_1, \cdots, x_n; \epsilon]$ of T_0 has the property that $T(x_1), \cdots, T(x_n)$ are linearly independent.

3. On some closure formulas. We begin with the finite topology. For the notions of irreducibility and density see Jacobson [5].

3. 1. THEOREM. *Let G be an additive group in $A(\mathfrak{X}, \mathfrak{Y})$. If $R_l(G)$ is a two-fold transitive ring of transformations from $\mathfrak{N}(G)$ to $\mathfrak{N}(G)$ then the closure of G in the finite topology is the set of all T in $A(\mathfrak{X}, \mathfrak{Y})$ such that $\mathfrak{M}(T) \supset \mathfrak{M}(G)$ and $\mathfrak{N}(T) \subset \mathfrak{N}(G)$.*

By Theorem 2. 1, the closure of G in the finite topology is contained in the set of this theorem. To show that every transformation of the set is in the closure in the finite topology it is sufficient to show that if $x_1, \cdots, x_n$ are any n vectors in $\mathfrak{X}$, linearly independent modulo $\mathfrak{M}(G)$ and if $y_1, \cdots, y_n$ are any n vectors in $\mathfrak{N}(G)$, then there exists a transformation $T \varepsilon G$ such that $T(x_i) = y_i$, $i = 1, \cdots, n$.

We proceed by induction. Let $x \not\varepsilon \mathfrak{M}(G)$, $y \varepsilon \mathfrak{N}(G)$. Then there exists $T \varepsilon G$ such that $T(x) \neq 0$. As y and $T(x)$ are in $\mathfrak{N}(G)$, by hypothesis there exists $U \varepsilon R_l(G)$ such that $U[T(x)] = y$. Since $UT \varepsilon G$, the proof for $n = 1$ is complete. For $n = 2$ we consider x_1 and x_2 in $\mathfrak{X}$ any two vectors linearly independent modulo $\mathfrak{M}(G)$ and y_1 and y_2 in $\mathfrak{N}(G)$. By the case $n = 1$ and the fact that G is an additive group there will exist $T \varepsilon G$ such that $T(x_i) = y_i$, $i = 1, 2$, if we can always find a $T \varepsilon G$ such that $T(x_1) = 0$ while $T(x_2) \neq 0$. We suppose that no such T exists in G. By the two-fold transitive nature of $R_l(G)$, this implies that for every $T \varepsilon G$, $T(x_1)$ and $T(x_2)$ are linearly dependent. As shown by Jacobson [5, p. 232] the mapping $y = T(x_1) \rightarrow T(x_2) = V(y)$ is single-valued and defined everywhere in $\mathfrak{N}(G)$. Furthermore, since G is an additive group and one-fold transitive, the domain $\mathfrak{N}(G)$ of this mapping is closed under addition and V is additive on $\mathfrak{N}(G)$. Moreover $\mathfrak{N}(G)$ has the property that if $y \varepsilon \mathfrak{N}(G)$, then $\lambda y \varepsilon \mathfrak{N}(G)$ for each scalar λ. Thus $\mathfrak{N}(G)$ is a linear manifold. We show that V is homogeneous on $\mathfrak{N}(G)$. For every $y \varepsilon \mathfrak{N}(G)$, there exist scalars α and β not both zero such that $\alpha y + \beta V(y) = 0$. If $\beta \neq 0$, then $V(y)$ is a multiple of y while if $\beta = 0$, then as $\alpha \neq 0$ we have $y = 0$ and again $V(y)$ is a multiple of y. Given $y = T(x_1) \neq 0$ and a scalar $\alpha \neq 0$, we have a transformation $U \varepsilon R_l(G)$ such that $UT(x_1) = \alpha y$. If $T(x_2) = 0$, then $V(y) = V(\alpha y) = \alpha V(y) = 0$. If $T(x_2) \neq 0$, then also $T(x_1) \neq 0$ and $T(x_2) = \mu T(x_1)$, $\mu \neq 0$. Since $UT(x_1) = \alpha T(x_1)$, it follows that $UT(x_2) = \alpha T(x_2)$. Thus $V(\alpha y) = UT(x_2) = \alpha V(y)$. To see that $V(y)$ is always the same multiple of y we need now consider only linearly independent y's. If $V(y_1) = \alpha_1 y_1$, $V(y_2) = \alpha_2 y_2$ where y_1 and y_2 are linearly independent in $\mathfrak{N}(G)$, then $V(y_1 + y_2) = \alpha_3(y_1 + y_2) = \alpha_1 y_1 + \alpha_2 y_2$ and $\alpha_1 = \alpha_2 = \alpha_3$. Thus $V(y) = \alpha y$ for some scalar α and therefore $T(x_2 - \alpha x_1) = 0$ for every $T \varepsilon G$. This contradicts the fact that x_1 and x_2 were chosen linearly independent modulo $\mathfrak{M}(G)$.

The induction argument for $n > 2$ proceeds along the lines of the analogous portion of the induction used by Jacobson [5, Theorem 6]. We suppose that the theorem is true for $n - 1$. As above, it is sufficient to show that for $x_1, \cdots, x_n$ linearly independent modulo $\mathfrak{M}(G)$ there exists a transformation $T \varepsilon G$ such that $T(x_1) = \cdots = T(x_{n-1}) = 0$, $T(x_n) \neq 0$. Consider the subgroup G' of G of all $T \varepsilon G$ which vanish on $x_1, \cdots, x_{n-2}$.

$\mathfrak{M}(G')$ contains $\mathfrak{M}(G)$ and $x_1, \cdots, x_{n-2}$ but, by induction hypothesis, no non-trivial linear combination of x_{n-1} and x_n. Thus x_{n-1} and x_n are linearly independent modulo $\mathfrak{M}(G')$. Then since $R_l(G') \supseteq R_l(G)$, we may use the first part of the proof to assert the existence of a transformation $T \varepsilon G'$ such that $T(x_{n-1}) = 0$, $T(x_n) \neq 0$. This completes the induction.

In the course of the proof we showed the following.

3.2. COROLLARY. *Under the hypotheses of Theorem* 3.1, $\mathfrak{N}(G)$ *is a linear manifold.*

3.3. COROLLARY. *Under the hypotheses of Theorem* 3.1 *the closure of* G *in the strong and weak topologies is the set of all* $T \varepsilon A(\mathfrak{X}, \mathfrak{Y})$ *such that* $\mathfrak{M}(T) \supset \mathfrak{M}(G)$ and $\mathfrak{N}(T) \subset \overline{\mathfrak{N}(G)}$.

This follows immediately from Theorem 3.1, Corollary 3.2 and Theorem 2.2.

3.4. THEOREM. *Let* G *be an additive group in* $A(\mathfrak{X}, \mathfrak{Y})$. *If* $R_r(G)$ *is a two-fold transitive ring of transformations from* $\mathfrak{X}/\mathfrak{M}(G)$ *to itself then the closure of* G *in the finite topology is the set of all* $T \varepsilon A(\mathfrak{X}, \mathfrak{Y})$ *such that* $\mathfrak{M}(T) \supset \mathfrak{M}(G)$ *and* $\mathfrak{N}(T) \subset \mathfrak{N}(G)$.

As in Theorem 3.1 it is sufficient to show that if $x_1, \cdots, x_n$ are any n vectors of $\mathfrak{X}$ linearly independent modulo $\mathfrak{M}(G)$ and if $y_1, \cdots, y_n$ are in $\mathfrak{N}(G)$ then there exists a transformation $T \varepsilon G$ such that $T(x_i) = y_i$, $i = 1, \cdots, n$. By the definition of $\mathfrak{N}(G)$ there exist vectors z_i in $\mathfrak{X}$ and transformations $U_i \varepsilon G$ such that $U_i(z_i) = y_i$, $i = 1, \cdots, n$. As shown by Jacobson [5, Theorem 6], $R_r(G)$ is a dense ring of transformations from $\mathfrak{X}/\mathfrak{M}(G)$ to itself. Using the notation of section 2, where $R_r(G)$ is defined, we see that there exist transformations V_i in $R_r(G)$ such that $V_i(x_i') = z_i'$ and $V_i(x_j') = 0'$ for $j \neq i$, $i, j = 1, \cdots, n$. Let T_i be any transformation in $R_2(G)$ such that $T_i'' = V_i$, $i = 1, \cdots, n$. Then $\sum_{i=1}^{n} U_i T_i \varepsilon G$ and as each U_i vanishes on $\mathfrak{M}(G)$,

$$\sum_{i=1}^{n} U_i T_i(x_j) = U_j T_j(x_j) = U_j(z_j) = y_j.$$

This completes the proof.

We note that under the hypothesis of the theorem, $\mathfrak{N}(G)$ is necessarily a linear manifold. For let $\sum_{i=1}^{n} \lambda_i y_i$ be a linear combination of vectors in $\mathfrak{N}(G)$. Then using the notation of the proof,

$$\sum_{i=1}^{n} U_i T_i\Big(\sum_{j=1}^{n} \lambda_j x_j\Big) = \sum_{i=1}^{n} \lambda_i y_i .$$

Thus, as in the case of Corollary 3. 3, we have the following result.

3. 5. COROLLARY. *Under the hypotheses of Theorem* 3. 4, *the closure of* G *in the strong and weak topologies is the set of all* $T \in A(\mathfrak{X}, \mathfrak{Y})$ *such that* $\mathfrak{M}(T) \supset \mathfrak{M}(G)$ *and* $\mathfrak{N}(T) \subset \overline{\mathfrak{N}(G)}$.

We can express the results of Corollaries 3. 3 and 3. 5 in another way by introducing the concept of annihilator.

3. 6. DEFINITION. *Let* $B \subset A(\mathfrak{X}, \mathfrak{Y})$. *By the left (right) annihilator of* B *we mean the set* $B^l (B^r)$ *of all* $U \in A(\mathfrak{Y}, \mathfrak{X})$ *such that* $UT = 0$ $(TU = 0)$ *for every* $T \in B$.

The following is readily verified.

3. 7. LEMMA. B^l *is the set of all* $U \in A(\mathfrak{Y}, \mathfrak{X})$ *such that* $U(y) = 0$ *for every* $y \in \overline{\mathfrak{N}_1(B)}$. B^r *is the set of all* $U \in A(\mathfrak{Y}, \mathfrak{X})$ *such that* $\mathfrak{N}(U) \subset \mathfrak{M}(B)$.

If we set $B^{lr} = (B^l)^r$, $B^{rl} = (B^r)^l$ then, by Lemma 3. 7, $B^{lr} = \{T \in A(\mathfrak{X}, \mathfrak{Y}) \mid \mathfrak{N}(T) \subset \overline{\mathfrak{N}_1(B)}\}$, $B^{rl} = \{T \in A(\mathfrak{X}, \mathfrak{Y}) \mid \mathfrak{M}(T) \supset \mathfrak{M}(B)\}$. Thus we have the following result.

3. 8. THEOREM. *Under the conditions of Corollary* 3. 3 *or Corollary* 3. 5, *the closure of the additive group* G *in either the strong or weak topology is the set* $G^{rl} \cap G^{lr}$.

We show next that if $\mathfrak{Y}$ is an infinite dimensional Banach space there is always an additive group G in $A(\mathfrak{X}, \mathfrak{Y})$ satisfying the conditions of Corollaries 3. 3 and 3. 5 such that the formula of Theorem 3. 8 properly contains the closure of G in the finite topology. This follows from the fact that $\mathfrak{Y}$ must always contain a submanifold which is not closed. This may be seen as follows. There exists a sequence $\{y_n\}$ of vectors in $\mathfrak{Y}$ such that no finite subset is linearly dependent. Let $\mathfrak{N}$ be the linear manifold in $\mathfrak{Y}$ generated by $\{y_n\}$. If $\mathfrak{N}$ is closed then $\{y_n\}$ is a Hamel basis for the Banach space $\mathfrak{N}$. But Mackey [9, p. 159] has shown that the power of the Hamel basis for an infinite-dimensional Banach space is at least that of the continuum. Hence $\mathfrak{N}$ is not closed. The additive group $G = \{T \in A(\mathfrak{X}, \mathfrak{Y}) \mid \mathfrak{N}(T) \subset \mathfrak{N}\}$ is closed in the finite topology by Theorem 3. 1, say, but has the larger closure $G^{rl} \cap G^{lr}$ in the strong and weak topologies.

4. On linear manifolds closed in the uniform topology. We now relate the property of a linear manifold being closed in the uniform topology of $A(\mathfrak{X}, \mathfrak{Y})$ to its having the maximum possible closure in the strong and weak topologies. We begin with a study of irreducible algebras in $A(\mathfrak{X}, \mathfrak{X})$. By an algebra is meant here a set which is both a linear manifold and a ring. The term *dense* is defined in 1.

4.1. THEOREM. *An irreducible algebra $\mathfrak{A}$ in $A(\mathfrak{X}, \mathfrak{X})$, closed in the uniform topology, is dense in $A(\mathfrak{X}, \mathfrak{X})$.*

Let us suppose that $\mathfrak{A}$ is not two-fold transitive. This gives us (Jacobson [5, p. 232]) two linearly independent elements of $\mathfrak{X}$, x_1 and x_2, such that for no $T \varepsilon \mathfrak{A}$ do we have $T(x_1) = 0$ and $T(x_2) \neq 0$. As shown by Jacobson this defines an additive mapping $V\colon T(x_1) \rightarrow T(x_2)$ of $\mathfrak{X}$ into itself which commutes with every $T \varepsilon \mathfrak{A}$. Since $\mathfrak{A}$ is a linear manifold, $V(\lambda x) = \lambda V(x)$ for each scalar λ. Now since $VT = TV$ for every $T \varepsilon \mathfrak{A}$ it is impossible that $V(x) = 0$ for $x \neq 0$, for otherwise we can choose T such that $T(x) = y$ where $V(y) \neq 0$, arriving at a contradiction. Next we consider the mappings W_i of $\mathfrak{A}$ into $\mathfrak{X}$ defined by

$$W_i(T) = T(x_i) \qquad i = 1, 2.$$

Each W_i is a homogeneous and additive transformation defined on the Banach space $\mathfrak{A}$ (in the uniform topology) and is bounded since

$$\| W_i(T) \| = \| T(x_i) \| \leqq \| x_i \| \, \| T \| .$$

By the irreducibility of $\mathfrak{A}$ each of these transformations has $\mathfrak{X}$ as its range. Since V is $1-1$, $W_1(T) = 0$ if and only if $W_2(T) = 0$. We let $\mathfrak{M} \subset \mathfrak{A}$ be the closed linear manifold for which $W_1(T) = W_2(T) = 0$. The quotient space $\mathfrak{A}/\mathfrak{M}$ is a Banach space with the norm of each coset L given by $\| L \| = \inf \| T \|$, $T \varepsilon L$. We define transformations $W_i'(L)$ on $\mathfrak{A}/\mathfrak{M}$ to $\mathfrak{X}$ by the rule

$$W_i'(L) = W_i(T), T \varepsilon L, \qquad i = 1, 2.$$

Each W_i' is linear and bounded for

$$\| W_i'(L) \| = \| W_i(T) \| \leqq \| W_i \| \, \| T \|$$

for each $T \varepsilon L$, hence $\| W_i'(L) \| \leqq \| W_i \| \, \| L \|$. In fact it may be shown that $\| W_i' \| = \| W_i \|$. By a well-known theorem of Banach [2, p. 41], W_i' is an isomorphism between $\mathfrak{A}/\mathfrak{M}$ and $\mathfrak{X}$ and $W_i'^{-1}$ is bounded. From the relation $W_2'W_1'^{-1}W_1' = W_2'$ it may be seen that $W_2'W_1'^{-1}(x) = V(x)$ for all $x \varepsilon \mathfrak{X}$. Consequently V is continuous. Since $\mathfrak{A}$ is irreducible, by a lemma of Bochner and Phillips [3, p. 412] it follows that V is a multiple of the identity transformation. Thus there is a scalar λ such that $V(x)=\lambda x$, whence $T(x_2-\lambda x_1)=0$ for every $T \varepsilon \mathfrak{A}$. Again by the irreducibility of $\mathfrak{A}$, $x_2-\lambda x_1=0$, which contradicts the fact that x_1 and x_2 were chosen to be linearly independent. Thus $\mathfrak{A}$ is two-fold transitive and by the results of Jacobson [5, p. 233], $\mathfrak{A}$ is dense.

4.2. Corollary. *An irreducible algebra $\mathfrak{A}$ in $A(\mathfrak{X},\mathfrak{X})$ is dense in the strong and weak topologies.*

The closure $\bar{\mathfrak{A}}$ of $\mathfrak{A}$ in the uniform topology is an irreducible algebra. $\bar{\mathfrak{A}}$ is contained in the closure of $\mathfrak{A}$ in the strong and weak topologies. Consequently the closure of $\bar{\mathfrak{A}}$ in the strong and weak topologies is the same as the closure of $\mathfrak{A}$ in these topologies. By Theorem 4.1, $\bar{\mathfrak{A}}$ is dense, hence dense in the strong and weak topologies.

4.3. Corollary. *If B is a linear manifold in $A(\mathfrak{X},\mathfrak{Y})$ which is closed in the uniform topology when considered as a linear manifold in $A(\mathfrak{X}/\mathfrak{M}(B),\mathfrak{Y})$ and if $R_r(B)$ is irreducible, then the closure of B in the finite topology is the set of all $T \varepsilon A(\mathfrak{X},\mathfrak{Y})$ such that $\mathfrak{M}(T)\supset\mathfrak{M}(B)$ and $\mathfrak{R}(T)\subset\mathfrak{R}(B)$.*

The assumption on closure implies that $R_r(B)$ is closed. It is also readily verified that $R_r(B)$ is a linear manifold and therefore an irreducible algebra. The conclusion then follows from Theorems 4.1 and 3.4.

4.4. Corollary. *If B is a linear manifold in $A(\mathfrak{X},\mathfrak{Y})$ with $\mathfrak{R}(B)$ closed in $\mathfrak{Y}$, B closed in the uniform topology and $R_l(B)$ irreducible, then the conclusion of Corollary 4.3 holds for the finite, weak, strong and weakly finite topologies.*

We show first that $\mathfrak{R}(B)$ is a linear manifold in $\mathfrak{Y}$. If $y_i \varepsilon \mathfrak{R}(B)$, $x_0 \varepsilon \mathfrak{X}$, $x_0 \bar{\varepsilon} \mathfrak{M}(B)$, then there exists a transformation $T_i \varepsilon B$ such that $T_i(x_0)\neq 0$, and thus by hypothesis such that $T_i(x_0)=y_i$, $i=1,2$. Then $(T_1+T_2)(x)=y_1+y_2$ and $\mathfrak{R}(B)$ is a linear manifold and therefore a Banach space. By the closure of B, $R_l(B)$ is closed in the uniform topology of $A(\mathfrak{R}(B),\mathfrak{R}(B))$. By Theorems 4.1 and 3.1 the closure in the finite topology is the desired set. Since $\mathfrak{R}(B)$ is a closed linear manifold, by Theorem 2.1 this is also the closure in the strong, weak and weakly finite topologies.

5. Ideals and quotient spaces in $A(\mathfrak{X}, \mathfrak{Y})$. By analogy we formulate the following definition.

5. 1. DEFINITION. *An additive group G in $A(\mathfrak{X}, \mathfrak{Y})$ is called a left (right) ideal in $A(\mathfrak{X}, \mathfrak{Y})$ if $R_1(G) = A(Y, Y)$ $(R_2(G) = A(X, X))$.*

5. 2. THEOREM. *If G is a left ideal in $A(\mathfrak{X}, \mathfrak{Y})$ its closure in the finite, strong, weak and weakly finite topologies is the set G^{rl}. If G is a right ideal its closure in the finite topology is the set of all $T \varepsilon A(\mathfrak{X}, \mathfrak{Y})$ such that $\mathfrak{N}(T) \subset \mathfrak{N}(G)$. Its closure in the strong and weak topologies is the set G^{lr}.*

This follows readily from the result of section **3**. It is then seen that a left (right) ideal is weakly closed if and only if it is the right (left) annihilator of some set in $A(\mathfrak{Y}, \mathfrak{X})$. From this we conclude the following results.

5. 3. COROLLARY. *The lattice of closed linear subspaces of $\mathfrak{X}(\mathfrak{Y})$ is isomorphic to the lattice of left (right) annihilators in $A(\mathfrak{X}, \mathfrak{Y})$.*

For the case $\mathfrak{X} = \mathfrak{Y}$ this was given for right annihilators by Kakutani and Mackey [8, p. 53].

5. 4. COROLLARY. *Let B be any subset of $A(\mathfrak{X}, \mathfrak{Y})$. Then B^{rl} (B^{lr}) is the smallest weakly closed left (right) ideal in $A(\mathfrak{X}, \mathfrak{Y})$ which contains B.*

It is clear that B^{rl} (B^{lr}) is a weakly closed left (right) ideal in $A(\mathfrak{X}, \mathfrak{Y})$ containing B. On the other hand, if E is a weakly closed left ideal containing B, then $E^r \subset B^r$, $E^{rl} \supset B^{rl}$ and $E = E^{rl}$. Thus $E \supset B^{rl}$.

5. 5. COROLLARY. *For a left ideal G in $A(\mathfrak{X}, \mathfrak{Y})$ the following statements are equivalent.*

(1) *G is one-fold transitive.*

(2) *G is dense in the finite topology.*

(3) *G is dense in the weakly finite topology.*

(4) *G is dense in the strong topology.*

(5) *G is dense in the weak topology.*

That (1) implies (2) follows from Theorem 3. 1. Only the implication (5) → (1) need be shown. Suppose that G is not one-fold transitive. Then there exists $x_0 \neq 0$ in $\mathfrak{X}$ and $y_0 \varepsilon \mathfrak{Y}$ such that $T(x_0) = y_0$ is impossible for $T \varepsilon G$. By the left ideal property, $T(x_0) = 0$ for every $T \varepsilon G$. Let $T_0 \varepsilon A(\mathfrak{X}, \mathfrak{Y})$ have the property that $T_0(x_0) = y \neq 0$ and suppose $y^* \varepsilon \mathfrak{Y}^*$ has the property

$y^*(y) = 1$. Then no $T \varepsilon G$ is in the weak neighborhood $N[T_0; x_0; y^*; \frac{1}{2}]$ which is contradictory to (5).

We consider next the left ideal $A_l(\mathfrak{M})$ of $A(\mathfrak{X}, \mathfrak{Y})$ where $A_l(\mathfrak{M}) = \{T \varepsilon A(\mathfrak{X}, \mathfrak{Y}) \mid T(x) = 0$ for every $x \varepsilon \mathfrak{M}\}$ and $\mathfrak{M}$ is a closed linear manifold in $\mathfrak{X}$, and $A_r(\mathfrak{N}) = \{T \varepsilon A(\mathfrak{X}, \mathfrak{Y}) \mid T(x) \varepsilon \mathfrak{N}$ for every $x \varepsilon \mathfrak{X}\}$ where $\mathfrak{N}$ is a linear manifold in $\mathfrak{Y}$. In particular we investigate the difference (quotient) spaces $A_l(\mathfrak{M})/A_l(\mathfrak{M}) \cap A_r(\mathfrak{N})$ and $A_r(\mathfrak{N})/A_l(\mathfrak{M}) \cap A_r(\mathfrak{N})$.

5.6. THEOREM. *Let* $\mathfrak{M}$ *and* $\mathfrak{N}$ *be closed linear manifolds in* $\mathfrak{X}$ *and* $\mathfrak{Y}$ *respectively. The quotient space* $A_l(\mathfrak{M})/A_l(\mathfrak{M}) \cap A_r(\mathfrak{N})$ *possesses a linear continuous* $1 - 1$ *image which is dense in the finite topology of* $A(\mathfrak{X}/\mathfrak{M}, \mathfrak{Y}/\mathfrak{N})$. *If there exists a projection of* $\mathfrak{Y}$ *on* $\mathfrak{N}$ *the image is all of* $A(\mathfrak{X}/\mathfrak{M}, \mathfrak{Y}/\mathfrak{N})$ *and the correspondence is an isomorphism.*

Let C be the quotient space in question. Since $A_l(\mathfrak{M})$ and $A_r(\mathfrak{N})$ are closed in $A(\mathfrak{X}, \mathfrak{Y})$ in the uniform topology, C, with the customary definition of norm for quotient spaces, is a Banach Space. Let T_1 and T_2 be two transformations in $A_l(\mathfrak{M})$. They belong to the same coset in C if and only if the range of $T_1 - T_2$ is in $\mathfrak{N}$; i. e., if $T_1(x)$ and $T_2(x)$ are always in the same coset of $\mathfrak{Y}$ modulo $\mathfrak{N}$. Now for any $T \varepsilon A_l(\mathfrak{M})$, $T(x)$ has the same value over any coset of $\mathfrak{X}$ modulo $\mathfrak{M}$. To every coset $L \varepsilon C$ there corresponds a transformation T' from $\mathfrak{X}/\mathfrak{M}$ to $\mathfrak{Y}/\mathfrak{N}$ defined as follows. Let $x' \varepsilon \mathfrak{X}/\mathfrak{M}$, $L \varepsilon C$, T be a representative of L and x a representative of x' Then we take $T'(x')$ as the coset of $\mathfrak{Y}/\mathfrak{N}$ which contains $T(x)$. The correspondence $L \to T'$ is readily shown to be additive and homogeneous. Since

$$\| T'(x') \| = \inf_{y \varepsilon T'(x')} \| y \| \leqq \inf_{x \varepsilon x'} \| T(x) \| \leqq \inf_{x \varepsilon x'} \| T \| \, \| x \|$$
$$= \| T \| \, \| x' \| ,$$

where T is any representative of L we see that $\| T' \| \leqq \| L \|$, that $T' \varepsilon A(\mathfrak{X}/\mathfrak{M}, \mathfrak{Y}/\mathfrak{N})$, and that the mapping is continuous. It is $1 - 1$, for if $T' = 0$ is the image of L then for every $T \varepsilon L$, $T(x) \varepsilon \mathfrak{N}$ for every x and $T \varepsilon A_r(\mathfrak{N})$, whence L is the zero of C.

To see that the image is dense we consider n linearly independent elements $x_1', \cdots, x_n'$ of $\mathfrak{X}/\mathfrak{M}$ and any n elements $y_1', \cdots, y_n'$ of $\mathfrak{Y}/\mathfrak{N}$. If $x_1, \cdots, x_n$; $y_1, \cdots, y_n$ are representatives of these cosets and we choose $T \varepsilon A_l(\mathfrak{M})$ such that $T(x_i) = y_i$, $i = 1, \cdots, n$, then the image T' of the coset containing T has the property that $T'(x_i') = y_i'$, $i = 1, \cdots, n$.

Next suppose that there exists a projection of $\mathfrak{Y}$ on $\mathfrak{N}$. Let $W \varepsilon A(\mathfrak{X}/\mathfrak{M}, \mathfrak{Y}/\mathfrak{N})$. Consider W_1 defined from $\mathfrak{X}$ to $\mathfrak{Y}/\mathfrak{N}$ by $W_1(x) = W(x')$ where

$x \varepsilon x'$, $x' \varepsilon \mathfrak{X}/\mathfrak{M}$. The author has shown [11] that the existence of a projection of $\mathfrak{Y}$ on $\mathfrak{N}$ is equivalent to the existence of a continuous linear transformation R with domain $\mathfrak{Y}/\mathfrak{N}$ and range $\mathfrak{Y}$ such that the image under R of each coset is an element in that coset. Since W_1 vanishes on $\mathfrak{M}$, $RW_1 \varepsilon A_l(\mathfrak{M})$. Let L be the coset of C which contains RW_1. Since under our correspondence we have $L \rightarrow W$, the correspondence has all of $A(\mathfrak{X}/\mathfrak{M}, \mathfrak{Y}/\mathfrak{N})$ as its range. By a theorem of Banach [2, p. 41], the correspondence is an isomorphism.

5.7. THEOREM. *In the notation of Theorem* 5.6, $A_r(\mathfrak{N})/A_r(\mathfrak{N}) \cap A_l(\mathfrak{M})$ *possesses a linear continuous* $1-1$ *image which is dense in the finite topology of* $A(\mathfrak{M}, \mathfrak{N})$. *If there exists a projection of* $\mathfrak{X}$ *on* $\mathfrak{M}$, *the image is all of* $A(\mathfrak{M}, \mathfrak{N})$ *and the correspondence is an isomorphism.*

If L is a coset of $A_r(\mathfrak{N})$ modulo $A_r(\mathfrak{N}) \cap A_l(\mathfrak{M})$, then T_1 and T_2 of $A_r(\mathfrak{N})$ are in L if and only if $T_1 - T_2$ vanishes on $\mathfrak{M}$. Thus to L we can correspond a transformation T' which is defined on $\mathfrak{M}$ and has range in $\mathfrak{N}$. If $T \varepsilon L$, then T' is its contraction to the domain $\mathfrak{M}$. It is readily seen that the correspondence is additive and homogeneous. Furthermore since $\| T' \| \leqq \| T \|$ for every representative $T \varepsilon L$, $\| T' \| \leqq \| L \|$. Thus the mapping is continuous. As in Theorem 5.6, we see that the image is dense in the finite topology of $A(\mathfrak{M}, \mathfrak{N})$.

If $T' \varepsilon A(\mathfrak{M}, \mathfrak{N})$ and if there is a projection P of $\mathfrak{X}$ on $\mathfrak{M}$, then $T'P \varepsilon A_r(\mathfrak{N})$ and T' is the image of the coset containing $T'P$ under our correspondence. Again by Banach's Theorem [2, p. 41] the correspondence is an isomorphism in this case.

6. On the notion of centralizer.

Jacobson [5] has used the notion of centralizer effectively in his study of the density of rings of transformations. We can carry this notion over to additive groups in $A(\mathfrak{X}, \mathfrak{Y})$ and obtain partial success in the direction of Jacobson's results for rings.

6.1. DEFINITION. *If* U *and* V *are additive transformations defined on* $\mathfrak{X}$ *to* $\mathfrak{X}$ *and on* $\mathfrak{Y}$ *to* $\mathfrak{Y}$ *respectively, then* (U, V) *is said to be in the centralizer of* $B \subset A(\mathfrak{X}, \mathfrak{Y})$ *if* $UT = TV$ *for every* $T \varepsilon B$.

6.2. THEOREM. *Let* G *be an additive group in* $A(\mathfrak{X}, \mathfrak{Y})$ *where* $R_1(G)$ *and* $R_2(G)$ *are one-fold transitive. If for every* (U, V) *in the centralizer of* G *at least one of the transformations* U *and* V *is a multiple of the identity transformation, then* G *is two-fold transitive.*

We again use the argument of Jacobson [5, Theorem 6]. G is certainly one-fold transitive. If G is not two-fold transitive there exist elements x_1 and x_2, linearly independent in $\mathfrak{X}$, such that $T(x_1) = 0$ implies $T(x_2) = 0$ for $T \varepsilon G$. This defines a mapping U: $T(x_1) \to T(x_2)$ of $\mathfrak{Y}$ into itself. It must also be impossible to find $S \varepsilon R_2(G)$ such that $S(x_1) = 0$ while $S(x_2) \neq 0$. Thus there exists a mapping V: $S(x_1) \to S(x_2)$ of $\mathfrak{X}$ into itself. U and V are additive.

Let $T \varepsilon G$, $x \varepsilon \mathfrak{X}$. There exists a transformation $S \varepsilon R_2(G)$ such that $S(x_1) = x$. Then $TV(x) = TVS(x_1) = TS(x_2)$. Also $UT(x) = UTS(x_1)$. But since $S \varepsilon R_2(G)$, $TS \varepsilon G$. Thus $UTS(x_1) = TS(x_2)$ and $UT = TV$ for every $T \varepsilon G$. If V is a multiple of the identity, $V(x) = \lambda x$, then $x_2 - \lambda x_1 = 0$, which is impossible, and the same argument holds if U is a multiple of the identity.

We turn our attention to the case $\mathfrak{X} = \mathfrak{Y}$. Here the centralizer of a set B is the set of all endomorphisms U of $\mathfrak{X}$ such that $UT = TU$ for every $T \varepsilon B$.

6.3. THEOREM. *Any non-zero ideal B (left or right) in $A(\mathfrak{X}, \mathfrak{X})$ has as its centralizer the set of multiples of the identity transformation.*

Let B be a non-zero left ideal. Arnold [1, p. 30] has shown that B possesses a minimal left ideal consisting of all transformations of the form $x^*(x)y$ where x^* is a fixed linear functional defined on $\mathfrak{X}$ and y ranges over $\mathfrak{X}$. Let U be in the centralizer of B. Then

$$x^*[U(x)]y = U[x^*(x)y] \tag{1}$$

for every $x \varepsilon \mathfrak{X}$, $y \varepsilon \mathfrak{X}$. Since $x^* \neq 0$ we can choose $x \varepsilon \mathfrak{X}$ such that $x^*(x) = 1$. If we then take $y = \lambda z$, where λ is a complex scalar, we see from (1) that $\lambda U(z) = U(\lambda z)$ so that U is homogeneous. Also $U(y)$ is a multiple of y for each y. An argument used in the proof of Theorem 3.1 shows that U is a multiple of the identity transformation.

Let B be a non-zero right ideal. Arnold [1, p. 31] has shown that B possesses a minimal right ideal consisting of all transformations of the form $x^*(x)y$ where y is a fixed vector in $\mathfrak{X}$ and x^* ranges over $\mathfrak{X}^*$. If U is in the centralizer of B then (1) holds for every $x \varepsilon \mathfrak{X}$ and $x^* \varepsilon \mathfrak{X}^*$. Taking $x^*(x) = 1$ and $z^* = \lambda x^*$ we see from (1) $U(\lambda y) = \lambda U(y)$ which yields homogeneity for multiples of y. Hence (1) can be written in the form

$$x^*[U(x)]y = x^*(x)U(y). \tag{2}$$

This shows that $U(y)$ is a multiple of y, say $U(y) = \alpha y$. Then from (2) we have

$$x^*[U(x) - \alpha x]y = 0$$

for every $x^* \varepsilon \mathfrak{X}^*$, $x \varepsilon \mathfrak{X}$. Also $y \neq 0$. Therefore $U(x) = \alpha x$ for each x and thus U is a multiple of the identity transformation.

This theorem together with Jacobson's theorem [5, Theorem 6] shows that a one-fold transitive left ideal is dense, a fact which is contained in Corollary 5. 5 above.

YALE UNIVERSITY
AND
CORNELL UNIVERSITY.

BIBLIOGRAPHY

1. B. H. Arnold, "Rings of operators on vector spaces," *Annals of Mathematics*, vol. 45 (1944), pp. 24-29.
2. S. Banach, *Theorie des opérations linéaires*, Warsaw, 1932.
3. S. Bochner and R. S. Phillips, "Absolutely convergent Fourier expansions for non-commutative normed rings," *Annals of Mathematics*, vol. 43 (1942), pp. 409-418.
4. I. Gelfand, "Normierte Ringe," *Recueil Mathématique (Mat. Sbornik)*, New Series, vol. 9 (1941), pp. 3-23.
5. N. Jacobson, "Structure theory of simple rings without finiteness assumptions," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 228-245.
6. ———, "The radical and semi-simplicity for arbitrary rings," *American Journal of Mathematics*, vol. 67 (1935), pp. 300-320.
7. ———, "On the theory of primitive rings," *Annals of Mathematics*, vol. 48 (1947), pp. 8-21.
8. S. Kakutani and G. W. Mackey, "Two characterizations of real Hilbert space," *Annals of Mathematics*, vol. 45 (1944), pp. 50-58.
9. G. W. Mackey, "On infinite dimensional linear spaces," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 155-207.
10. J. von Neumann, "Zur Algebra der Funktionaloperatoren und Theorie der normalen Operatoren," *Mathematische Annalen*, vol. 102 (1929), pp. 370-427.
11. B. Yood, "Transformations between Banach spaces in the uniform topology," *Annals of Mathematics*, vol. 50 (1949), pp. 486-503.

INVARIANTS OF INTERSECTION OF CERTAIN PAIRS OF CURVES IN N-DIMENSIONAL SPACE.*

By Chuan-Chih Hsiung.

Introduction. It is well known that there is a projective tac-invariant of Mehmke-Smith [5, 6, 10][1] associated with two plane curves having ordinary contact at a nonsingular point. Several authors [3, 7, 8, 9, 11, 12] have extended this result to two curves C, C' in n-dimensional space having the same tangent at an ordinary point 0 by imposing different conditions on their respective osculating linear spaces at the point 0. In particular, B. Segre [8] considered the most general case in which the two curves C, C' have at the point 0 the same osculating linear spaces of dimensions $1, \cdots, r$, where r is any fixed integer satisfying $1 \leqq r \leqq n-1$. The purpose of this paper is to study the situation in which the two curves C, C' have distinct tangents at the common point 0. This investigation may be regarded as a generalization of some results for ordinary space given by the author in a forthcoming paper [4].

CHAPTER I. Two Curves Intersecting at an Ordinary Point With Distinct Osculating Linear Spaces.

1. Derivation of Invariants.

Let two curves C, C' in n-dimensional projective space S_n $(n \geqq 3)$ intersect at an ordinary point 0 with distinct osculating linear spaces S_k, S'_k $(k = 1, \cdots, n-1)$ respectively. Let $x_1, \cdots, x_n$ represent projective nonhomogeneous coordinates of a point in the space S_n. If we choose the point 0 to be the origin and the x_i-axis to be the line of intersection of the osculating linear spaces S_i and S'_{n-i+1}, then the power series expansions of the two curves in the neighborhood of the point 0 may be written in the form

$$C: \quad x_i = a_i x_1^{\,i} + \cdots \qquad (i = 2, \cdots, n), \tag{1}$$

$$C': \quad x_i = b_i x_n^{\,n-i+1} + \cdots \qquad (i = 1, \cdots, n-1). \tag{2}$$

---

* Received September 1, 1948; presented to the American Mathematical Society, September 10, 1948.

[1] Numbers in brackets refer to the bibliography at the end of the paper.

In order to find projective invariants of the two curves C, C' at the point 0, we consider the most general projective transformation of coordinates which leaves invariant the point 0 and all the osculating linear spaces of the curves C, C' at the point 0. This transformation is expressed in terms of the nonhomogeneous coordinates by the equations

$$x_i = \alpha_{ii} x_i^* / D \qquad (i = 1, \cdots, n), \tag{3}$$

where

$$D = 1 + \sum_{k=1}^{n} \alpha_{0k} x_k^*, \tag{4}$$

and the α's are arbitrary. The effect of this transformation on equations (1), (2) is to produce two other systems of equations of the same form whose coefficients, indicated by stars, are given by the formulas

$$\begin{aligned} \alpha_{ii} a_i^* &= \alpha_{11}{}^i a_i & (i = 2, \cdots, n), \\ \alpha_{ii} b_i^* &= \alpha_{nn}{}^{n-i+1} b_i & (i = 1, \cdots, n-1). \end{aligned} \tag{5}$$

Eliminating the α's from equations (5), we see that *the expressions*

$$I_i = b_1^{n-i} / a_n^{i-1} (a_i / b_i)^{n-1} \qquad (i = 2, \cdots, n-1) \tag{6}$$

are projective invariants associated with the point 0 *of intersection of the curves* C, C'.

2. Metric and projective characterizations of a general invariant I_i.

For the purpose of finding a simple metric characterization of a general invariant I_i, we make a projective transformation which leaves the point 0 unchanged and carries the x_1-, $\cdots$, x_n-axes into n mutually perpendicular new axes. Let Γ, Γ' be the transformed curves of C, C', and $\bar{x}_1, \cdots, \bar{x}_n$ be the nonhomogeneous Cartesian coordinates of a point in the space S_n referred to the new orthogonal coordinate system; then the power series expansions of the curves Γ, Γ' in the neighborhood of the point 0 may be written in the form

$$\Gamma\colon \ \bar{x}_i = \bar{a}_i \bar{x}_1^{\,i} + \cdots \qquad (i = 2, \cdots, n), \tag{7}$$

$$\Gamma'\colon \ \bar{x}_i = \bar{b}_i \bar{x}_n^{\,n-i+1} + \cdots \qquad (i = 1, \cdots, n-1), \tag{8}$$

and the invariant I_i takes the form

$$I_i = \bar{b}_1^{n-i} / \bar{a}^{i-1} (\bar{a}_i / \bar{b}_i)^{n-1} \qquad (i = 2, \cdots, n-1). \tag{9}$$

Let $\bar{\rho}_i$, $\bar{\sigma}_i$ $(i=1,\cdots,n-1)$ be respectively the i-th curvatures of the curves Γ, Γ' at the point 0, then from the generalized Frenet-Serret formulas [2]

$$(10)\quad \bar{a}_i=\bar{\rho}_1\bar{\rho}_2\cdots\bar{\rho}_{i-1}/i!,\qquad b_i=\bar{\sigma}_1\bar{\sigma}_2\cdots\bar{\sigma}_{n-i}/(n-i+1)!\qquad (i=1,\cdots,n-1),$$

and therefore

$$(11)\quad I_i\sim[(\bar{\rho}_1\cdots\bar{\rho}_{i-1})^{n-1}/(\bar{\rho}_1\cdots\bar{\rho}_{n-1})^{i-1}][(\bar{\sigma}_1\cdots\bar{\sigma}_{n-1})^{n-i}/(\bar{\sigma}_1\cdots\bar{\sigma}_{n-i})^{n-1}]\qquad (i=2,\cdots,n-1),$$

where $\sim$ denotes that both its members are equal except for a constant factor. Thus we arrive at the following conclusion.

Let ρ_i, σ_i $(i=1,\cdots,n-1)$ *be respectively the i-th curvatures of the two curves* C, C' *at the point* 0, *then the quantities*

$$(12)\quad J_i=(\rho_{i-1}\sigma_{n-i+1})^{n-2}/[(\rho_1\cdots\rho_{i-2}\rho_i\cdots\rho_{n-1})(\sigma_1\cdots\sigma_{n-i}\sigma_{n-i+2}\cdots\sigma_{n-1})]\qquad (i=2,\cdots,n-1)$$

are projective invariants, and

$$(13)\quad J_2\sim I_2,\qquad J_i\sim I_i/I_{i-1},\qquad I_i\sim J_2J_3\cdots J_i\quad (i=3,\cdots,n-1).$$

Let the "point at infinity" on the x_i-axis $(i=1,\cdots,n)$ be denoted by 0_i, and Γ_i, Γ'_i be the projections of the curves C, C' from the center $0_2\cdots 0_{i-1}0_{i+1}\cdots 0_{n-1}$ onto the space $00_10_i0_n$ $(i=2,\cdots,n-1)$. In order to interpret a general invariant I_i projectively, we consider a system of cones of order $n-1$ and vertex 0 in the space $00_10_i0_n$ such that the polar surfaces of orders $1,\cdots,n-i-1$, $n-i+1,\cdots,n-1$ of any point in the plane 00_i0_n with respect to any cone of the system (the last of these polars being the cone itself) all pass through the line 00_1. The equations of this system of cones can easily be written as

$$(14)\qquad x_2=\cdots=x_{i-1}=x_{i+1}=\cdots=x_{n-1}=0,$$

$$(15)\qquad x_1^{n-i}\sum\beta_{jk}x_i^jx_n^k+\sum\gamma_{hl}x_i^hx_n^l=0,$$

$$\text{with } j,k=0,1,\cdots,i-1;\ j+k=i-1, \text{ and}$$
$$h,l=0,1,\cdots,n-1;\ h+l=n-1$$

where the β's and γ's are arbitrary. If the cones of this system have contact of the $i(n-1)$-th order with the curve Γ_i at the point 0, then their equation (15) reduces to

---

[2] Cf. [1, p. 16] or [7, p. 396].

$$a_i^{n-1}x_1^{n-i}x_n^{i-1} - a_n^{i-1}x_i^{n-1} + \sum A_{jk}x_i^jx_n^k = 0, \tag{16}$$

with $j = 0, 1, \cdots, n-2$; $k = 1, \cdots, n-1$; $j + k = n-1$,

where the A's are arbitrary. Among the cones (16) we can determine a unique one which passes through the tangent of the curve Γ'_i at the point 0 and with respect to which the polar surfaces of orders $1, \cdots, n-2$ of any point in the plane 00_10_n all pass through the line 00_i. It is easily seen that for this cone all A's vanish and equation (16) then becomes

$$a_i^{n-1}x_1^{n-i}x_n^{i-1} - a_n^{i-1}x_i^{n-1} = 0. \tag{17}$$

Similarly, we may have a cone of order $n-1$ and vertex 0 with the equation

$$b_i^{n-1}x_1^{n-i}x_n^{i-1} - b_1^{n-i}x_i^{n-1} = 0, \tag{18}$$

which has contact of the $(n-1)(n-i+1)$-th order with the curve Γ'_i at the point 0 and has the same other properties as that of the cone (17). The two cones (17), (18) determine a pencil of cones of order $n-1$ and vertex 0 in the space $00_10_i0_n$. Belonging to this pencil there are two degenerate cones K_1 consisting of the plane 00_10_n, counted $n-1$ times, and K_2 consisting of the plane 00_i0_n, counted $n-i$ times, and the plane 00_10_i, counted $i-1$ times. Hence *the invariant I_i is equal to the cross ratio of the four cones K_1, K_2,* (17), (18).

CHAPTER II. Two Curves Intersecting at an Ordinary Point With Distinct Tangents But Certain Common Osculating Linear Spaces.

3. Derivation of invariants. Let two curves C, C' in n-dimensional projective space $S_n \equiv S'_n$ intersect at an ordinary point 0 with tangents t, t' and osculating linear spaces S_k, S'_k of dimensions $k (k = 2, \cdots, n-1)$ respectively. In this chapter we shall consider the case in which the tangents t, t' are distinct but $S_2 \equiv S'_2, \cdots, S_r \equiv S'_r$, where r is any fixed integer satisfying $2 \leqq r \leqq n-1$, and there is no other relation among the osculating linear spaces of the two curves C, C' at the point 0.

We first introduce in the space S_n any system of nonhomogeneous projective coordinates $x_1, \cdots, x_n$ with origin at the point 0, having the x_1-, x_2-axes as the tangents t, t' respectively, the x_3-, $\cdots$, x_r-axes as independent lines lying in the common osculating linear spaces of dimensions $3, \cdots, r$ respectively, and the x_{r+1}-, $\cdots$, x_n-axes as independent lines not lying in the common osculating linear space S_r but lying in the $(r+1)$-dimensional spaces of intersection of S_{r+1} and S'_n, S_{r+2} and $S'_{n-1}, \cdots, S_n$ and S'_{r+1} respectively. Then the power series expansions of the two curves in the neighborhood of the point 0 may be written in the form

$$(19) \qquad C\colon\ x_i = a_i x_2^i + \cdots \qquad (i = 2, \cdots, n),$$

$$(20) \qquad C'\colon \begin{array}{ll} x_1 = b_2 x_2^2 + \cdots, \quad x_i = b_i x_2^i + \cdots & (i = 3, \cdots, r), \\ x_{r+j} = b_{r+j} x_2^{n-j+1} + \cdots & (j = 1, \cdots, n-r). \end{array}$$

Let us now make a most general projective transformation of coordinates which leaves the point 0 and the tangents t, t' invariant, and changes the other axes in such a way that the new axes have the same properties as the old. This transformation is expressed in terms of the nonhomogeneous coordinates by the equations

$$(21) \qquad \begin{array}{ll} x_1 = 1/D(\alpha_{11} x_i^* + \sum_{k=3}^{n} \alpha_{1k} x_k^*), & \\ x_i = 1/\mathrm{D} \sum_{k=i}^{n} \alpha_{ik} x_k^* & (i = 2, \cdots, r), \\ x_j = \alpha_{jj} x_j^*/D & (j = r+1, \cdots, n), \end{array}$$

where D is defined by expression (4) and the α's are arbitrary. The effect of this transformation on equations (19), (20) is to produce two other systems of equations of the same form whose coefficients, indicated by stars, are given by the formulas

$$(22) \qquad \alpha_{ii} a_i^* = \alpha_{11}{}^i a_i \qquad (i = 2, \cdots, n),$$

$$(23) \qquad \begin{array}{ll} \alpha_{11} b_2^* = \alpha_{22}{}^2 b_2, \quad \alpha_{ii} b_i^* = \alpha_{22}{}^i b_i & (i = 3, \cdots, r), \\ \alpha_{jj} b_j^* = \alpha_{22}{}^{n+r+1-j} b_j & (j = r+1, \cdots, n). \end{array}$$

Eliminating the α's from equations (22), (23), we see that *the expressions*

$$(24) \qquad I_i = (a_i/b_i)^3 (b_2/a_2)^i \qquad (i = 3, \cdots, r),$$

$$(25) \qquad J_i = (a_i/b_i)^3 b_2{}^{2(n+r+1)-3i}/a_2{}^{3i-n-r-1} \qquad (i = r+1, \cdots, n)$$

are projective invariants associated with the point 0 *of intersection of the curves* C, C'.

4. Metric and projective characterizations of general invariants I_i and J_i. Let ρ_i, σ_i $(i=1,\cdots,n-1)$ be respectively the i-th curvatures of the two curves C, C' at the point 0, then it is easy to show that *the quantities*

$$(26)\qquad \begin{aligned} &H_3=\rho_2/\sigma_2, \quad H_i=(\rho_{i-1}/\sigma_{i-1})^3\sigma_1/\rho_1 && (i=4,\cdots,r),\\ &K_{r+1}=(\rho_r/\sigma_r\cdots\sigma_{n-1})^3\ \sigma_1^{2n-2r-1}/\rho_1^{r+2-n}, &&\\ &K_i=(\rho_{i-1}\sigma_{n+r+1-i})/(\rho_1\sigma_1) && (i=r+2,\cdots,n) \end{aligned}$$

are projective invariants, and

$$(27)\qquad \begin{aligned} &H_3^3=I_3, \quad H_i=I_i/I_{i-1}, \quad I_i=H_3^3H_4\cdots H_i \qquad (i=4,\cdots,r),\\ &K_{r+1}\sim J_{r+1}/I_r, \quad K_i^3\sim J_i/J_{i-1}, \quad J_i\sim I_rK_{r+1}K_{r+2}^3\cdots K_i^3\\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (i=r+2,\cdots,n). \end{aligned}$$

Let the "point at infinity" on the x_i-axis $(i=1,\cdots,n)$ be denoted by 0_i, and Γ_i, Γ'_i be the projections of the curves C, C' from the center $0_3\cdots 0_{i-1}0_{i+1}\cdots 0_n$ onto the space $00_10_20_i$ $(i=3,\cdots,r)$. It is immediately seen that the equations of the curves Γ_i, Γ'_i are

$$(28)\qquad x_3=\cdots=x_{i-1}=x_{i+1}=\cdots=x_n=0,$$

$$(29)\qquad \Gamma_i:\ x_2=a_2x_1^2+\cdots, \qquad x_i=a_ix_1^i+\cdots,$$

$$(30)\qquad \Gamma'_i:\ x_1=b_2x_2^2+\cdots, \qquad x_i=b_ix_2^i+\cdots.$$

Now we project Γ_i, Γ'_i onto the plane 00_10_i from a point V arbitrarily chosen in the plane 00_10_2. If the coordinates of the point V satisfy equations (28) and $x_1=\bar{x}_1$, $x_2=\bar{x}_2$, $x_i=0$, then the two projections are easily found to be given by the equations $X_2=X_3=\cdots=X_{i-1}=X_{i+1}=\cdots=X_n=0$ and

$$(31)\qquad X_i=a_iX_1^i+\cdots,$$

$$(32)\qquad X_i=b_i(-\bar{x}_2/\bar{x}_1)^iX_1^i+\cdots.$$

The cones projecting the curves Γ_i, Γ'_i from the point V have contact of order at least i along the line $0V$ if, and only if, the invariant of contact of the curves (31), (32) at the point 0 equals unit; then, and only then, the center

V of projection lies on the principal lines [2, p. 25] at the point 0 of the two curves Γ_i, Γ'_i. Hence these principal lines have equations (28), $x_i = 0$ and

$$a_i x_1^i + (-1)^{i+1} b_i x_2^i = 0. \tag{33}$$

On the other hand, let $Q(Q')$ be any three-point quadric of the curve Γ_i (Γ'_i) at the point 0 in the three-dimensional space $00_1 0_2 0_i$ satisfying the condition that the tangent plane of $Q(Q')$ at the other point $T'(T)$ of intersection of $Q(Q')$ with the tangent $t'(t)$ passes through the point $T(T')$. Then by means of expansions (29), (30) it is easy to show that the lines joining the point 0 to the other three points of intersection of the two conics determined by the quadrics Q, Q' and the plane (t, t') have equations (28), $x_i = 0$ and

$$a_2 x_1^3 - b_2 x_2^3 = 0. \tag{34}$$

If u be any one of the principal lines (33) and v be any one of the three lines (34), then *the projective invariant* I_i $(i = 3, \cdots, r)$ *associated with the point* 0 *of intersection of the curves* C, C' *is, except for sign, equal to the* $3i$-*th power of the cross ratio* (tt', uv).

In order to find a projectively geometric characterization of a general invariant J_i of expressions (25), we consider the projections Γ_i, Γ'_i of the curves C, C' from the center $0_3 \cdots 0_{i-1} 0_{i+1} \cdots 0_n$ onto the space $00_1 0_2 0_i$ $(i = r + 1, \cdots, n)$. These projections are readily found to be given by equations (28), (29) and

$$\Gamma'_i: \ x_1 = b_2 x_2^2 + \cdots, \qquad x_i = b_i x_2^{n+r+1-i} + \cdots. \tag{35}$$

As in **2**, we consider in the space $00_1 0_2 0_i$ a system of cones of order $i - 1$ and vertex 0, having contact of order $2(i - 1)$ with the curve Γ_i at the point 0 and satisfying the conditions that the polar surfaces of orders $1, \cdots, i - 3$, $i - 1$ of any point in the plane $00_2 0_i$ with respect to any cone of the system all pass through the line 00_1. The equations of this system are easily found to be (28) and

$$a_i x_2^{i-1} - a_2^{i-1} x_1^{i-2} x_i + \sum E_{lm} x_2^l x_i^m = 0, \tag{36}$$

$$\text{with } l = 0, 1, \cdots, i - 2; \ m = 1, \cdots, i - 1; \ l + m = i - 1,$$

where the E's are arbitrary. Similarly, we have in the space $00_1 0_2 0_i$ another system of cones of order $n + r - i$ and vertex 0, having contact of order

$2(n+r-i)$ with the curve Γ'_i at the point 0 and satisfying the conditions that the polar surfaces of orders $1, \cdots, n+r-i-2$, $n+r-i$ of any point in the plane 00_10_i with respect to any cone of the system all pass through the line 00_2. The equations of this system are (28) and

$$b_i x_1^{n+r-i} - b_2^{n+r-1} x_2^{n+r-i-1} x_i + \sum F_{lm} x_1^l x_j^m = 0, \tag{37}$$

$$\text{with} \quad l = 0, 1, \cdots, n+r-i-1; \; m = 1, \cdots, n+r-i;$$
$$l + m = n + r - i,$$

where the F's are arbitrary. If we further impose on the cones (36), (37) the conditions that the planes through the line 00_1, in which their intersection lies, reduce to a single plane counted $n+r-2$ times, then these planes coincide with one or other of the $n+r-2$ planes satisfying equations (28) and

$$a_i^{n+r-i} b_i^{i-2} x_2^{n+r-2} - a_2^{(i-1)(n+r-i)} b_2^{(i-2)(n+r-i)} x_i^{n+r-2} = 0. \tag{38}$$

On the other hand, if we impose on the cones (36), (37) the conditions that the planes through the line 00_2, in which their intersection lies, reduce to a single plane counted $n+r-2$ times, then these planes coincide with one or other of the $n+r-2$ planes satisfying equations (28) and

$$a_i^{n+r-i-1} b_i^{i-1} x_1^{n+r-2} - a_2^{(i-1)(n+r-i-1)} b_2^{(i-1)(n+r-i)} x_i^{n+r-2} = 0. \tag{39}$$

The line 00_i and the lines of intersection of the two pencils (38), (39) of planes determine in the space $00_10_20_i$ another pencil of planes, which intersect the plane 00_10_2 in $n+r-2$ lines whose equations are (28), $x_i = 0$ and

$$a_2^{i-1} b_i x_1^{n+r-2} - a_i b_2^{n+r-i} x_2^{n+r-2} = 0. \tag{40}$$

Let v be any one of the lines (34) and w be any one of the lines (40) in the plane 00_10_2, then *the cross ratio of the four lines* t, t', v, w *is equal to* $J_i^{1/[3(n+r-2)]}$.

MICHIGAN STATE COLLEGE,
EAST LANSING, MICH.

BIBLIOGRAPHY

1. C. Guichard, "Les courbes de l'espace à n dimensions," *Mémorial des Sciences Mathématiques*, Paris, vol. 29 (1928).

2. G. H. Halphen, "Sur les invariants différentiels des courbes gauches," *Journal de l'École Polytechnique,* vol. 28 (1880).
3. C. C. Hsiung, "Projective invariants of contact of two curves in space of n dimensions," *Quarterly Journal of Mathematics,* Oxford Series, vol. 17 (1946), pp. 39-45.
4. ———, "Invariants of intersection of certain pairs of space curves," to appear in the *Bulletin of the American Mathematical Society.*
5. R. Mehmke, "Einige Sätze über die räumliche Collineation und affinität welche sich auf die Krümmung von Curven und Flächen beziehen," *Schlömilchs Zeitschrift für Mathematik und Physik,* vol. 36 (1891), pp. 56-60.
6. ———, "Über zwei die Krümmung von Curven und das Gauss'sche Krümmungsmass von Flächen betreffende charakteristische Eigenschaften der linearen Punkttransformationen," *ibid.,* vol. 36 (1891), pp. 206-213.
7. B. Segre, "Sugli elementi curvilinei che hanno comuni le origini ed i relativi spazi osculatori," *Rendiconti dei Lincei,* (6), vol. 22 (1935), pp. 392-399.
8. ———, "On tac-invariants of two curves in a projective space," *Quarterly Journal of Mathematics,* Oxford Series, vol. 17 (1946), pp. 35-38.
9. C. Segre, "Sugli elementi curvilinei, che hanno comuni la tangente e il piano osculatore," *Rendiconti dei Lincei,* (5), vol. 33 (1924), pp. 325-329.
10. H. J. S. Smith, "On the focal properties of homographic figures," *Proceedings of the London Mathematical Society,* (1), vol. 2 (1869), pp. 196-248.
11. B. Su, "Note on a theorem of B. Segre," *Science Record, Academia Sinia,* vol. 1 (1942), pp. 16-19.
12. ———, "On certain tac-invariants of two curves in a projective space," *Quarterly Journal of Mathematics,* Oxford Series, vol. 17 (1946), pp. 116-118.

A CHARACTERIZATION OF NON-ISENTROPIC IRROTATIONAL FLOWS.* †

By DAVID GILBARG.

1. Introduction. The standard theory of irrotational gas flows postulates that the flows be isentropic, that is, that the entropy be constant throughout the region of flow. In this work, the isentropic condition is replaced by the weaker adiabatic hypothesis that entropy is constant along the individual streamlines. An apparently larger class of irrotational flows is thus obtained. Our main result is the characterization *in the large,* as well as in the small, of a wide class of these irrotational flows. It is shown that all steady plane subsonic irrotational flows are either isentropic or are vortex flows, (Theorem 2), and that if certain general conditions are placed on the flow boundaries, the same result is true for supersonic and mixed flows, (Theorems 4 and 6). Similar results are established for axially symmetric flows, (Theorems 3, 5). These results signify for the general theory that the class of adiabatic irrotational flows is not essentially wider than the isentropic flows, and that the isentropic assumption is generally superfluous. Hicks [1], Truesdell and Prim [2], and others, also have studied the relation between irrotationality and entropy variation, but from a different viewpoint than is taken here.

The proofs make essential use of uniqueness theorems for the Cauchy initial value problem, applied in this case to the differential equation for the velocity potential of isentropic irrotational flow. The discussion shows that more general uniqueness theorems than are now available would strengthen some of the results obtained here.

2. Non-isentropic irrotational flows in the small. Let the flow region be an arbitrary open connected set R in 3-space. A steady adiabatic irrotational flow of an ideal compressible fluid will be considered defined by functions u, v, w, p, ρ, η which are continuously differentiable in R with respect to space coordinates x, y, z, and satisfy the following equations:

$$(1) \qquad (V \cdot \nabla) V = -(\operatorname{grad} p)/\rho, \quad V = (u, v, w)$$

$$(2) \qquad \operatorname{div}(\rho V) = 0$$

---

* Received February 16, 1949.

† Prepared under Navy Contract No. N6onr-180, Task Order 5, with Indiana University.

(3) $$\operatorname{curl} V = 0$$

(4) $$p = f(\eta)\rho^{\gamma},\ f,\ f' \neq 0,\ \gamma = \text{constant} \neq 0$$

(5) $$d\eta/dt = V \cdot \operatorname{grad} \eta = 0.$$

The term 'irrotational flow' will be used here to mean flows of this type. Equations (1) and (2) are the usual equations of motion and continuity, $V = (u, v, w)$ being the velocity vector of the flow, ρ the density (assumed $\neq 0$), p the pressure; (3) is the condition of irrotationality; (4) is the equation of state for an ideal polytropic gas;[1] (5) is the statement that entropy η is constant for a particle, or in other words, that the medium is thermally non-conducting; this replaces the stronger *isentropic* condition, $\operatorname{grad} \eta = 0$, which is customarily assumed in the theory of irrotational gas flows.

We now consider the flow as divided into two sets: the *isentropic set* A is the set of points at which $\operatorname{grad} \eta = 0$; the *non-isentropic set* B is the set of points at which $\operatorname{grad} \eta \neq 0$. It is clear, by continuity, that A and B are respectively closed and open relative to R. It will be assumed that none of the components (i. e. maximal connected sets) of B consists entirely of stagnation points.

The vortex and helicoidal flows play a central part in what follows; we therefore fix our ideas with these definitions. An irrotational flow in a plane region G is a *vortex flow* if either (1) the velocity vector is constant in G, or (2) the streamlines are concentric circular arcs, and the flow speed q at any point in G a distance r from the common center is $q = k/r$, where k is constant for the flow. It follows that for vortex flows p and ρ are constant on streamlines, and either ρ or η may be arbitrary functions of the streamline. A *helicoidal flow* is a spatial irrotational flow obtained by normal superposition of a uniform flow on a vortex flow; that is, for suitable coordinates x, y, z, the velocity vector is given by $\operatorname{grad}(wz + k \arctan y/x)$, where w, k are constants for the flow. The streamlines are thus concentric helices of the same pitch. One sees that for helicoidal flows p and ρ are again constant on streamlines and that, if the flow is non-uniform, either p, ρ, or η may be an arbitrary function of $r (= (x^2 + y^2)^{\frac{1}{2}})$, whereas if the flow is uniform, either ρ or η may be an arbitrary function of the streamline, (but p is constant). The vortex flows are evidently special cases of the helicoidal flows, corresponding to flows for which either $w = 0$ or $k = 0$.

The vortex and helicoidal flows are irrotational and may be non-isentropic.

---

[1] The same results hold for more general equations of state; see concluding remarks.

It will be seen that, in the plane and space, respectively, they are essentially the only non-isentropic flows. This is first demonstrated *in the small* in the following theorem.

THEOREM 1. *The flow in any region of the non-isentropic set of an irrotational flow is helicoidal.*

Proof. From equations (1), (3), and (4), we have,

$$\tfrac{1}{2}\operatorname{grad}(u^2+v^2+w^2) = -(\operatorname{grad} p)/\rho = \rho^{\gamma-1}[f'\operatorname{grad}\eta + \gamma f(\operatorname{grad}\rho)/\rho].$$

The first equality implies that, in the neighborhood of any point at which $\operatorname{grad} p \neq 0$, $u^2+v^2+w^2 = q^2 = [q(p)]^2$ and $\rho = \rho(p)$. Hence, using both equalities, we obtain at every point of the flow,

$$0 = \operatorname{grad}\rho \times \operatorname{grad}\eta = \operatorname{grad} p \times \operatorname{grad}\eta = \operatorname{grad} q^2 \times \operatorname{grad}\eta.$$

It follows that in the neighborhood of every point of B, $\rho = \rho(\eta)$, $p = p(\eta)$, and $q = q(\eta)$, and, as a consequence, ρ, p, and q are constant on the streamline components of B. The equation of continuity (2) thus takes the simpler form in B, $\operatorname{div} V = 0$. In conjunction with (3), this establishes that the flow in any subregion $B^* \subset B$ is described by a velocity potential $\phi(x, y, z)$ which is a harmonic function, which, furthermore, satisfies the additional condition that $q = |\operatorname{grad}\phi|$ is constant on streamlines. Hamel [3] has shown that the velocity field of such a potential must be helicoidal.

As a particular case of the preceding theorem, we have,

COROLLARY 1. *The flow in any region of the non-isentropic set of a plane irrotational flow is a vortex flow.*

An independent proof of this result is obtained easily by function theoretic methods.

COROLLARY 2. *The flow in any region of the non-isentropic set of an axially symmetric irrotational flow is a uniform flow.*

This follows from the obvious fact that the only axially symmetric helicoidal flows are uniform.

The preceding result is essentially local in that it applies only to regions of the non-isentropic set. For a characterization of the non-isentropic irrotational flows in the large we must determine the influence of the various parts of the flow on one another. This is provided by the following lemmas.

LEMMA 1. *Any streamline of an irrotational flow lies entirely either in the isentropic set or in the non-isentropic set of the flow.*

Proof. Let S be a streamline containing a point P in B. Suppose, in contradiction to the theorem, that S meets A; then, since A is closed, there is a segment S' such that $P \in S' \subset S$, $S' \subset B$, and S' has an endpoint Q in A. Since the flow in the component of B containing S' is helicoidal, we have that on S', $|\operatorname{grad} \eta| = k \neq 0$, for some constant k. By continuity, $|\operatorname{grad} \eta| = k$ also at Q, contradicting the assumption that S' could have an endpoint in A. Therefore, S must lie entirely in B, and indeed in one component of B. This proves the lemma.

The preceding lemma shows that the sets A and B are invariant under the flow in the sense that particles in either set remain in that set as long as they are in the region of flow.

LEMMA 2. *Let $\dot{B}$ be the set of boundary points of the non-isentropic set; then if a point P of $\dot{B}$ is on a streamline C, C lies entirely in $\dot{B}$, and is a circular helix*[2] *which extends to the boundary of the flow region (perhaps at infinity); q, ρ, and p are constant on C.*

Proof. Let $C = x(t), y(t), z(t)$ be the streamline[3] through $P = x(t_0), y(t_0), z(t_0)$. In B there is a sequence of points P_n converging to P, and let $C_n = x_n(t), y_n(t), z_n(t)$ be the (helical) streamlines through P_n, the parametrization of C_n being chosen such that $P_n = (x_n(t_0), y_n(t_0), z_n(t_0))$. By virtue of the continuous dependence of the sequence $x_n(t), y_n(t), z_n(t)$ on the initial values at t_0, the sequence converges uniformly to $x(t)$, $y(t)$, $z(t)$, in every closed interval (t_0, t), and hence the limit curve C of the sequence of helices C_n is itself a helix and lies in $\dot{B}$. The functions q, p, ρ on C, being the limit of the corresponding quantities q_n, p_n, ρ_n on C_n must evidently be constant on C. C cannot contain a stagnation point, therefore does not terminate in R. This completes the proof.

We shall now limit the discussion to plane and axially symmetric flows.

3. Plane and axially symmetric flows in the large.

For the proof of the basic Lemma 3, we require certain uniqueness theorems from the theory of partial differential equations. We consider in particular the quasilinear equations for the velocity potential of an isentropic flow.

---

[2] 'Circular helix' is here used also for the degenerate cases of a circle and straight line. This convention of including the limiting forms of a curve under one heading will be adopted throughout, except where explicit reference is made to the limiting forms.

[3] The non-constant solutions $x(t)$, $y(t)$, $z(t)$, of the system, $dx/dt = u(x, y, z)$, $dy/dt = v(x, y, z)$, $dz/dt = w(x, y, z)$, are understood to define the streamlines of the flow.

Plane flows:

$$(6a)\quad (1-\phi_x{}^2/c^2)\phi_{xx}-2\phi_x\phi_y\phi_{xy}/c^2+(1-\phi_y{}^2/c^2)\phi_{yy}=0;$$

Axially symmetric flows:

$$(6b)\quad (1-\phi_x{}^2/c^2)\phi_{xx}-2\phi_x\phi_y\phi_{xy}/c^2+(1-\phi_y{}^2/c^2)\phi_{yy}+\phi_y/y=0;$$
$$c^2=c_0{}^2+(1-\gamma)/2(\phi_x{}^2+\phi_y{}^2),\qquad c_0=\text{constant}.$$

On a given smooth curve $C=x(s),\ y(s)$, where s is the arc length, ($y(s)>0$ for (6b)), let Cauchy data, $\phi=\phi(s)$; $\phi_x=u(s)$, $\phi_y=v(s)$, be given functions of s which satisfy the compatibility condition, $\phi'=ux'+vy'$. We consider solutions of (6a, b) having continuous second derivatives in a suitable neighborhood of C and on C. Then we can state the following uniqueness theorems:

𝔈—THE CASE OF ELLIPTIC DATA. *Let G be a one-sided neighborhood of C,[4] and let $\phi(x,y)$, $\bar{\phi}(x,y)$ be solutions of* (6a, b) *in G taking on the limit values ϕ, u, v on C, and satisfying in G and on C the inequality $\phi_x{}^2+\phi_y{}^2<c^2$ [i. e. $\phi_x{}^2+\phi_y{}^2<2c^2/(1+\gamma)$]; then $\phi(x,y)\equiv\bar{\phi}(x,y)$ in G.*[5]

ℌ—THE CASE OF HYPERBOLIC DATA. *Let C be non-characteristic with respect to the given data, and let $u^2+v^2>c^2$. Let $\phi(x,y)$ be a solution of* (6a, b) *taking on the limit values ϕ, u, v on C and satisfying the inequality $\phi_x{}^2+\phi_y{}^2>c^2$ in a characteristic triangle region G bounded by C and two characteristics of* (6a, b), *defined with respect to ϕ). Then if $\bar{\phi}(x,y)$ is any solution of* (6a, b) *in G satisfying the same conditions on C, $\phi(x,y)\equiv\bar{\phi}(x,y)$ in G* [5].

The corresponding theorem for the case of parabolic data on C seems to be incomplete at present. It is stated here in terms of the flow quantities under somewhat restricted hypotheses, and is proved in Section **4**.

𝔓—THE CASE OF PARABOLIC DATA. *On C, let the flow speed be sonic, the tangential component of velocity $\neq 0$, and the normal derivative of the speed $\neq 0$; (it follows that on one side of C there is a neighborhood on which any solution of* (6a, b) *with values ϕ, u, v on C is supersonic and on the other side a neighborhood in which the flow is subsonic). Let $\phi(x,y)$ be a solution*

---

[4] It is assumed that C is part of the boundary of G, and that some neighborhood of every interior point of C contains no boundary points of G other than points of C.

[5] For a proof see [4], in particular, pp. 325-326 of this reference, where a simple proof is given for the case that C is an analytic curve, and the given data on C are analytic, which is the case of interest for Lemma 3.

of (6a) *taking on the initial values* ϕ, u, v, *on* C *and defined either* (a) *in a characteristic triangle* G *on the supersonic side of* C, *or* (b), *in a subsonic neighborhood* G; *then if* $\bar{\phi}(x, y)$ *is any other solution defined in* G *and satisfying the same conditions on* C, $\phi(x, y) \equiv \bar{\phi}(x, y)$ *in* G.[6]

The proof of this theorem is deferred to Section 4. It is to be noted that Theorem 𝔓 does not apply to the simple situation that C is a straight streamline on which the velocity is sonic, for, in this case, the normal derivative of the speed is zero on C. This is the essential gap in the theory which follows. We formulate this conjectured theorem for reference purposes.

𝔓′. *Let* C *be a straight line segment of a streamline and on* C *let the speed be sonic and constant. Then in a neighborhood on either side of* C, *the only solution of* (6a, b) *is the uniform sonic flow.*

This uniqueness theorem, if true, will probably hold in any neighborhood of C contained within the strip bounded by the two perpendiculars to C at its end points.

The following lemma will prove to be essential in the sequel.

LEMMA 3. *In a plane irrotational flow, let* C *be either* (a) *a circular streamline segment, or* (b) *a closed circular streamline, or* (c) *an infinite straight streamline which does not approach the boundary of the flow. On* C *let the speed, pressure, and density be constant, and let* C *be curved if the speed is sonic on it. Then, respectively,* (a) *in some neighborhood containing every subarc of* C *in its interior, or* (b) *in a concentric annulus containing* C, *or* (c) *in an infinite parallel strip containing* C, *the flow is a vortex flow. If the flow is axially symmetric, and* C *is straight, then the flow in the respective neighborhoods of* C *is uniform.*

Proof. Consider first plane flows. (a) Through any interior point $a \in C$ consider the curve orthogonal to the streamline family. Let L be a segment of the curve such that the streamlines through L sweep out a full neighborhood G of C. There exists such a congruence of streamlines since $q \neq 0$ in the neighborhood of C. If the speed is sonic on C, C is curved by hypothesis, and G can be chosen so small that the streamlines in it have non-zero curvature.

L intersects the sets A and B in subsets A' and B' which are respectively closed and open on L. Since A' is closed, it may contain closed subarcs of L, and forms at all its other points a nowhere dense set. If a is interior to a subarc σ of A' or is an end point, then there is a subregion of G belonging to A swept out by the congruence of streamlines through this arc, and consequently equation (6a) for the potential of an isentropic flow describes the flow in this region. The vortex flow is an obvious solution of (6a) satisfying the stated conditions on C, and also the hypotheses of the appropriate uniqueness theorem $\mathfrak{E}$, $\mathfrak{H}$, or $\mathfrak{P}$.[7] It follows that according as a is an interior point or end point of σ, the vortex flow is the only solution in a one-sided neighborhood of C, or in a two-sided neighborhood containing every subarc of C in its interior. (If the flow is sonic or supersonic on C, repeated application of $\mathfrak{P}$ or $\mathfrak{H}$ is generally necessary to establish uniqueness in this neighborhood.) Also, if a is an isolated point of A', then clearly the flows on both sides of C are vortex flows (belonging to B), which, combined with C, form a connected vortex flow.

---

[6] For the case that G is on the supersonic side of C, this theorem has been proved by L. Bers [6]. The method of proof is quite different from the one shown in section 4 of this paper.

Suppose, then that a is neither an isolated nor an interior point of A'. If there are subarcs of A' in every neighborhood of a, let σ be such a segment, b the end point of σ closer to a, and b' the farther end point. The region $S \subset G$ swept out by the streamlines through σ belong to the isentropic set, and in it, therefore, equation (6a) applies. Since b is a boundary point of B', hence of B, the streamline $Q \ni b$ is by Lemma 2, a circular arc on which q, p, ρ are constant. Again an obvious solution of (6) which holds in S for these values on Q is that given by the vortex flow. If the vortex flow is wholly subsonic or supersonic in the interior of S, and if G was initially chosen sufficiently small,[8] we can then apply the appropriate uniqueness theorems $\mathfrak{E}$, $\mathfrak{H}$, or $\mathfrak{P}$, to establish the uniqueness of the vortex flow in all of S. If, however, neither of the inequalities $q^2 \gtrless \gamma p/\rho$ holds exclusively for the vortex flow in the interior of S, then the proof of uniqueness is in two steps, first by means of $\mathfrak{E}$ (or $\mathfrak{H}$) in the subsonic (or supersonic) strip between Q and the sonic circular streamline in S (which must exist as a consequence of the uniqueness proof), and then by application of theorem $\mathfrak{P}$ to the strip of S between the sonic streamline and the streamline through b'. It has thus been proved that the flow through any subarc $\sigma \subset A'$ is a vortex flow.

Since the flow in G through the subarcs of A' and B' consists of vortex flow strips, it remains only to show that these belong to a *single* vortex flow in G. That the streamlines through L are concentric circular arcs is evident. Then in G, $q = k(r)/r$, (or $q = k(r)$ if the streamlines are all straight), where $k(r)$ is the constant of the vortex flow at points a distance r from the common center (or from a common parallel). That $k'(r) \equiv 0$ follows, since $k'(r)$ is continuous and equal to zero on a dense set of values of r on L, namely, on the subarcs of A' and B'. Hence $k(r)$ is a constant and $q = k/r$ in G for fixed k. The constancy of p and ρ on streamlines in G is evident. This completes the proof of part (a).

---

[7] If C is an arc of radius r, $dq/dr = -q/r \neq 0$, so that the hypotheses of $\mathfrak{P}$ are satisfied if the speed is sonic on C.

[8] It suffices to choose G as follows: If $q^2 \gtrless \gamma p/\rho$ on C, let G be so small that the corresponding inequality holds throughout it; furthermore, if $q^2 \geq \gamma p/\rho$ in G, let M be an upper bound of $q/(\gamma p/\rho)^{\frac{1}{2}}$ in G, and consider the region cut out by the four curves through the end points of C which make the acute angle arc sin $1/M$ with C and with the streamlines in G; let this new region be the G of the proof.

Case (b) that C is a closed circle is reduced to (a) in an evident way by subdividing C into several overlapping arcs to which the lemma now applies. This determines an annulus about C in which the flow is a vortex flow.

To prove the lemma for case (c), we confine our attention to a strip on each side of C of width less than the distance of C from the boundary of the flow. Consider any finite segment $C' \subset C$, then by part (a) of the lemma, the flow in a neighborhood G of C' must be uniform. As in the proof of (a), we can define an orthogonal segment L and on it sets A' and B', where now the streamlines through the boundary points of B' are infinite straight lines on which q, p, ρ are constant, and $q^2 \neq \gamma p/\rho$ if G is sufficiently small. By paralleling the proof of part (a), one sees that it suffices to prove that, if q, p, ρ are constant with $q^2 \neq \gamma p/\rho$ on an infinite straight streamline Q, and the flow is uniform and isentropic in a finite strip s of width δ on one side of Q, then the flow is uniform in the infinite strip S of width δ. This follows by virtue of Lemma 1 and then $\mathfrak{E}$ or $\mathfrak{H}$.

The proof of the lemma for the case that the flow is axially symmetric and C is straight is the same except for evident verbal changes.

The restriction that C be curved in case the speed is sonic could be dropped if Theorem $\mathfrak{P}'$ were proved.

The proof of Lemma 3 can be greatly simplified if the data on the initial arc C are not sonic. This is seen as follows. Equations (6a, b) apply to the entire field of flow, and not merely to the isentropic regions, if c_0 is some appropriate function of the coordinates which is constant on streamlines. The generalized uniqueness Theorems $\mathfrak{E}$ and $\mathfrak{H}$ for this more general form of (6a, b) are contained implicitly in the Carleman uniqueness theorem [10]. With these stronger uniqueness theorems, it is clear that the proof of Lemma 3 is almost trivial for the case of nonsonic data on C.

The following theorems in the large can now be proved.

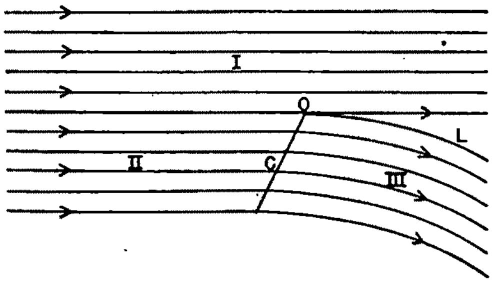
THEOREM 2. *A plane irrotational flow which is subsonic throughout a region R is either isentropic or a vortex flow in R.*

Proof. Assume the flow to be non-isentropic. Then by Theorem 1, Corollary 1, R contains a subregion of vortex flow $B^* \subset B$. Let $\bar{R}$ be the maximal vortex flow region in R containing B^*. We wish to show $\bar{R} = R$. If not, let P be a boundary point of $\bar{R}$. If $P \subset B$, then a neighborhood of it is a vortex flow region, which contains a subregion of $\bar{R}$, and is therefore a continuation of the vortex flow in $\bar{R}$, thereby contradicting the definition of $\bar{R}$ as a maximal region of vortex flow. Similarly, if P is a boundary of B, then the streamline through P is a circular arc on which q, p, ρ are constant. Therefore, by Lemma 3, a neighborhood of P is a vortex flow region, and, as above, this contradicts the maximal character of $\bar{R}$. Finally, if P is an interior point of A, then, in a neighborhood of P, equation (6a) holds for the velocity potential, the flow being isentropic in this neighborhood. The equation is also of elliptic type by hypothesis that the flow is subsonic. Since the flow is a vortex flow in a subregion of this neighborhood, then, by the well-known theorem on the analyticity of solutions of elliptic differential equations [7], the flow in the entire neighborhood is of vortex type. As in the previous two cases, this leads to a contradiction. It follows that $\bar{R}$, having no boundary points in R, is both open and closed in the connected set R; hence $R = \bar{R}$. This proves the theorem.

THEOREM 3. *An axially symmetric irrotational flow which is subsonic throughout a region R is either isentropic or a uniform flow in R.*

The proof proceeds as in Theorem 2, except for several obvious changes appropriate to axially symmetric flows.

One cannot expect to extend these theorems in all generality to mixed or supersonic flows, as the following simple counterexample shows.



In the figure let the flow in region I be uniform and non-isentropic, $\eta=\eta(y)$, a non-constant function. In II the flow is uniform, q having the same constant value as in I, but η, p, ρ are constant, and such that $q^2>\gamma p/\rho$. In region III, the flow is a non-uniform continuation of the flow past the characteristic C, the curve L being any suitable smooth curve forming a cusp at 0. The flow in the entire region I + II + III is thus neither isentropic nor a vortex flow.

A suitable generalization of Theorems 2 and 3 to mixed and supersonic flows requires limitations on the heretofore arbitrary character of the flow regions. The limitation we adopt here is one reflecting the physical condition that flows exist in the large, in the sense that particles do not enter and vanish through the boundaries of the flow region. This restriction is made precise by demanding that the flow region have *streamline boundaries*: that is, the boundary will consist of piecewise smooth curves which are streamlines of the flow on which we require, furthermore, that the velocity be continuous.

The above example shows that, even for these more restricted flow regions, extensions of Theorems 2 and 3 do not hold generally. However, appropriate generalizations are obtained if we exclude from the class of admissible boundary curves the following special types of curves:[9] (1) infinite half lines, (2) circular arcs, (3) closed curves, consisting in part of a circular arc, with cusps at both ends of the arc, (4) curves, consisting in part of a half line, with a cusp at the end of the half line. The boundary curve of the counter example flow shown in the figure is of type (4); similar examples can easily be constructed for flows with boundaries of the type (1)-(3). The admissible boundary curves obtained from the exclusion of types (1)-(4) will for the sake of brevity be called *non-cusped streamlines* (although this name signifies a smaller class than is being considered). With this definition, we have

THEOREM 4. *A plane subsonic or supersonic irrotational flow in a region bounded by non-cusped streamlines is either isentropic or a vortex flow.*

Proof. The subsonic case has already been proved for general regions in Theorem 2; we therefore confine ourselves to supersonic flows. Assume the flow is non-isentropic, and let B^* be a vortex flow component of B. The streamlines of B^* must all be closed circles or infinite straight lines, for otherwise (Lemma 1) a streamline $Q\subset B^*$ would terminate at the boundary of the flow region R, meeting there a curve which, at the same time would have to be a bounding streamline of B; this is possibly only if the latter

---

[9] These curves are assumed to be maximal, i. e., they are not contained within a larger connected portion of the boundary.

curve is a circular arc continuation of Q, and is therefore either a circular arc or an infinite half line, contrary to our assumption concerning boundary curves. Let $r(P)$ be the distance to any point P in the plane from the center of B^*, or, if the flow in B^* is uniform, from some fixed line parallel to the streamlines of B^*. Define m and M to be such that (1) the annulus $m < r(P) < M$ lies in the maximal vortex region containing B^*, and (2) m and M are respectively the minimum and maximum (possibly infinite) such values of m and M. Such an m and M exist since B^* itself is an annulus satisfying (1). Assume M is finite; we shall then show that the entire circle, or straight line, $C: r = M$, is a boundary streamline of the flow. If this were not so, then either it lies wholly interior to R, or consists only partly of boundary points of R. In the former case, the curve C is a streamline in R on which q, p, ρ are constant and $q^2 > \gamma p/\rho$, and hence to which Lemma 3b or 3c[10] can be applied. Thus there is a strip $M \leq r(P) < \bar{M}$ within which the flow is a vortex flow, thereby contradicting the choice of M as a maximum. On the other hand, C may consist partially of boundary points of R. These form a closed set C' on C, containing generally circular arcs and a set of points not dense in any segment of C. By definition of the class of boundary curves, the boundary curve containing one of these arcs must have a non-cusped vertex on C. It follows from the continuity of the velocity on the boundary that such a vertex is a stagnation point, which is impossible since it is on the boundary of a vortex flow. C' therefore does not contain arcs. But C' cannot contain other points of the flow boundary, for, as one readily sees, this would contradict the hypothesis that the boundary curve containing such a point must be a streamline on which the flow is continuous. It follows that either $M = \infty$, or the entire curve $r = M$ bounds the flow in R. In precisely the same way one shows that the curve $r = m$ is a bounding streamline. Since the flow region is connected, the curves $r = m$ and $r = M$ are the only boundary curves. Thus the flow is either isentropic or is a vortex flow in the annulus or parallel strip $m < r < M$.

By making suitable verbal changes in the preceding proof we have,

THEOREM 5. *An axially symmetric subsonic or supersonic irrotational flow in a region bounded by non-cusped streamlines is either isentropic or a uniform flow.*

Theorems 4 and 5 could be stated for mixed flows provided the conjectured uniqueness theorem $\mathfrak{P}'$ were true. For in this case, Lemma 3 would hold without exception, and hence the preceding proofs would go through just as well with the curve C a straight line on which $q^2 = \gamma p/\rho$. However, if C is curved, and $q^2 = \gamma p/\rho$ on it, the proof of Theorem 4 still holds without the limitation to supersonic or subsonic flows. To insure that C be curved, it suffices that the streamlines of the non-isentropic component B^* be curved. We thus have the following limited theorem for mixed flows.

---

[10] Lemma 3c requires that C does not approach the boundary of R; but if the boundary curve is a streamline, the proof of the lemma shows that the distance between C and this curve must stay above some fixed value greater than 0.

THEOREM 6. *If a plane irrotational flow in a region bounded by non-cusped streamlines is non-isentropic, and if the non-isentropic set contains a curved streamline, then the flow is a vortex flow.*

The proofs of preceding theorems assist in characterizing the wider class of flows obtained by dropping the limitation to non-cusped streamline boundaries. They show that the general irrotational flow with streamline boundary is "pieced together" from isentropic and vortex flows along circular arcs or straight lines in such a way that the new boundary curves formed by the piecing process are curves of the exceptional set (1)-(4). The flow shown in the figure is a simple example of this.

4. Uniqueness theorem for parabolic data. The proof of Theorem $\mathfrak{P}$ is achieved by the use of the well-known transformations of (6a) to the hodograph plane. Let us consider, for example, the Legendre transformation, $u = u(x, y)$, $v = v(x, y)$, $\Phi(u, v) = xu + yv - \phi$, by means of which (6a) is transformed into

$$(7)\qquad (1 - u^2/c^2)\Phi_{vv} + 2(uv/c^2)\Phi_{uv} + (1 - v^2/c^2)\Phi_{uu} = 0,$$
$$c^2 = c_0^2 + \tfrac{1}{2}(1 - \gamma)(u^2 + v^2), \qquad c_0 = \text{constant}.$$

To reduce the uniqueness theorem for (6a) to one for the above equation, it must first be established that any solution $\phi(x, y)$ of the former equation, which satisfies the hypothesis of Theorem $\mathfrak{P}$ on the initial curve C provides a transformation to the hodograph plane which is $1 - 1$ in the neighborhood of C. This is shown by Bers [6], who proves that C is mapped in a $1 - 1$ way onto an arc $\bar{C}$ of the circle $u^2 + v^2 = c^2$ $(= 2c_0^2/1 + \gamma)$ in the hodograph plane, and that the Jacobian of the transformation $\partial(u, v)/\partial(x, y)$ does not vanish on C, hence in a neighborhood.[11] Thus any solution $\phi(x, y)$ of (6a) which satisfies the prescribed conditions on C, is transformed into a solution $\Phi(u,v)$ of (7) which on $\bar{C}$ takes on the initial values $\Phi(s)=xu+yv'-\phi$, $\Phi_u=x(s)$, $\Phi_v=y(s)$. That such a solution is unique in a neighborhood of $\bar{C}$ is established by means of the following special case of a theorem due to Holmgren [8, 9].

---

[11] This is easily seen for the particular application of theorems $\mathfrak{P}$ to lemma 3, where C is a streamline. For one obtains, $|\partial(u, v)/\partial(x, y)| = (uv' - vu')^2/\phi'^2$ along any sonic initial curve C, and if, in particular, C is a streamline, on which, by hypothesis, $\partial q/\partial n \neq 0$ and $q = \phi' \neq 0$, then numerator and denominator of the right member are evidently non-zero.

Consider the equation,

$$L(w)=w_{xx}+aw_{xy}+bw_{yy}+cw_x+dw_y+ew=0,$$

where the coefficients a, b, c, d, e are analytic functions of x and y. Let C be a curve which is non-characteristic (at every point) with respect to $L(w)$, and let $w(x,y)$ be a solution which takes on the values $w=w_x=w_y=0$ on C, and which is continuous up to its second derivatives in a neighborhood and on C. Then $w(x,y)\equiv 0$ in a neighborhood of C.

Equation (7) satisfies the conditions of Holmgren's theorem, and thus $\Phi(u,v)$ is unique in a neighborhood of $\bar{C}$. This implies uniqueness of $\phi(x,y)$ in a neighborhood U of C. To complete the proof of Theorem $\mathfrak{P}$, it must be shown that if the solution $\phi(x,y)$ of the theorem in a characteristic triangle G on the supersonic side of C, then it is unique in all of G, and similarly for a neighborhood G on the subsonic side. However, this follows from the uniqueness result already proved for U, since in U, arbitrarily close to C, there are non-characteristic curves C', on which the initial data ϕ, ϕ_x, ϕ_y is uniquely determined, and to which the uniqueness theorems $\mathfrak{E}$ or $\mathfrak{H}$ can now be applied (according as C' is on the subsonic or supersonic side of C); this shows the uniqueness of ϕ in an arbitrarily large subregion of G, and therefore in all of G.

It is evident that Theorem $\mathfrak{P}'$ cannot be proved by this method, since in this case the hodograph transformation degenerates to a single point.

Remark 1. The restriction to the polytropic gas law (4) is not essential to the preceding results. One sees readily that the proof remain basically unaltered if the polytropic law is replaced by the general equation of state,

$$p=f(\rho,\eta), \qquad (f_\eta, f_\rho, f_{\rho\rho}\neq 0 \text{ for } \rho>0),$$

where $f(\rho,\eta)$ is analytic in ρ for every η. The latter condition is necessary to insure the analyticity of equations (6a, b); this is a requirement for the proof of $\mathfrak{P}$, and also for the theorem on the analytic character of solutions of elliptic equations, which is quoted in the proof of Theorems 2 and 3.

Remark 2. The preceding methods are not directly applicable to general spatial flows primarily because of difficulties in the proof of the generalized

Lemma 3. However, this obstacle could be easily overcome with the help of stronger uniqueness theorems than are now available for differential equations in three independent variables.

INDIANA UNIVERSITY.

REFERENCES.

[1] B. Hicks, "Diabatic flow of a compressible fluid," *Quarterly of Applied Mathematics*, vol. 6 (1948), pp. 221-237.

[2] C. Truesdell and R. Prim, *Vorticity and the thermodynamic state in the flow of an inviscid fluid*, Naval Ordnance Laboratory Memorandum 9416 (1947).

[3] G. Hamel, "Potentialströmungen mit konstanter Geschwindigkeit," *Sitzungsberichte der Preussichen Akademie der Wissenschaften*, 1937, pp. 5-20.

[4] H. Lewy, "Eindeutigkeit der Lösung der Anfangsproblems einer elliptischen Differentialgleichung zweiter Ordnung in zwei Veränderlichen," *Mathematische Annalen*, vol. 104 (1930-31), pp. 325-329.

[5] R. Courant and K. Friedrichs, *Supersonic flow and shock waves*, Interscience, 1948, Chapter 2.

[6] L. Bers, *On the continuation of a potential gas flow across the sonic line*, forthcoming NACA Technical Note.

[7] E. Hopf, "Über den funktionalen, insbesondere den analytischen Charakter der Lösungen elliptischer Differentialgleichungen zweiter Ordnung," *Mathematische Zeitschrift*, vol. 34 (1931-32), pp. 194-233.

[8] J. Hadamard, *Leçons sur la propagation des ondes*, Hermann, Paris, 1903, note I.

[9] E. Holmgren, "Über Systeme von linearen partieller Differentialgleichungen," *Öfversigt af Konigl. Svenska Vetenskaps-Akademiens Förhandlingar*, 1901, pp. 91-103.

[10] T. Carleman, "Sur un problème d'unicité pour les systèmes d'équations aux dérivées partielles à deux variables indépendantes," *Arkiv för Matematik, Astronomi och Fysik*, vol. 26B (1938-39), no. 17.

A CHARACTERIZATION OF THE NORMED VECTOR ORDERED SPACES OF CONTINUOUS FUNCTIONS OVER A COMPACT SPACE.*

By Leopoldo Nachbin.†

Given a compact space E, the system $\mathfrak{c}(E)$ of all real-valued continuous functions over E is a normed (real) vector ordered space, or more precisely, a normed vector lattice. The problem of finding a characterization up to an isomorphism of the normed vector sublattices of some $c(E)$ was considered previously by S. Kakutani, H. F. Bohnenblust, M. Krein and S. Krein.[1] In this note we shall consider the problem of giving a characterization up to an isomorphism of the normed vector ordered sub-spaces of some $c(E)$.[2]

Let S be a normed (real) vector space[3] which at the same time is an ordered vector space.[4] Following the usual terminology, we shall say that $x \in S$ is *positive* if $x \geqq 0$, and *negative* if $x \leqq 0$. We shall also say that x is *semi-positive* if $y \geqq y$ always implies $\| y \| \geqq \| x \|$, and that x is *semi-negative* if $y \leqq x$ implies $\| y \| \geqq \| x \|$.[5]

Our aim is to prove the following theorem.

Theorem 1. *In order that S be isomorphic in the order, vector and norm sense to a normed vector ordered sub-space of $c(E)$, for some compact space E, it is necessary and sufficient that* (1) *every element of S be either semi-positive or semi-negative, and* (2) *the set of all positive elements of S be closed.*

Before going into the proof of this theorem, we shall first state and prove the following theorem concerning the extension of linear functionals, which may be considered as a generalization of the well known Hahn-Banach theorem[6] to the case of ordered spaces, this last theorem being obtained as a particular case when the ordering relation is taken to be the discrete one.

---

* Received November 18, 1948.

† Fellow of the U. S. A. State Department.

[1] See the bibliography at the end of this paper.

[2] This case was also considered by M. Krein [8].

[3] See Banach [2].

[4] That is, S is a vector space which is (partially) ordered in such a way that $x \leqq y$ implies $x + z \leqq y + z$ and $\lambda x \leqq \lambda y$ for $\lambda \geqq 0$.

[5] The origin of this terminology is the fact that in the real line the notions of positiveness and semi-positiveness coincide.

THEOREM 2. *Let V be a vector sub-space of S, f be a linear functional defined over V and $k \geqq 0$. In order that there should exist a positive*[7] *continuous linear extension F of f defined over S such that $\| F \| \leqq k$ it is necessary and sufficient that $v \in V$, $w \in S$, $v \leqq w$ imply $f(v) \leqq k \| w \|$.*

The condition is clearly necessary because, if such an extension does exist and $v \in V$, $w \in S$, $v \leqq w$, then $f(v) = F(v) \leqq F(w) \leqq \| F \| \cdot \| w \| \leqq k \| w \|$.

To prove that the condition is sufficient let us assume that it is satisfied. For each $x \in S$, define $P(x)$ as the greatest lower bound of $\| y \|$ for all $y \in S$ such that $y \geqq x$. It is clear that $0 \leqq P(x)$, $P(\lambda x) = \lambda P(x)$ for $\lambda \geqq 0$ and $P(x_1 + x_2) \leqq P(x_1) + P(x_2)$. Our hypothesis that the condition is satisfied is equivalent to saying that $f(v) \leqq p(v)$ for all $v \in V$, where $p(x) = kP(x)$. Applying a known result,[8] we see that there exists a linear extension F of f, defined over S, such that $F(x) \leqq p(x)$ for all $x \in S$. If $x \geqq 0$, that is $-x \leqq 0$, then $P(-x) = 0$ and therefore $F(-x) \leqq p(-x) = 0$, that is $F(x) \geqq 0$, so that F is positive. Moreover it is clear from the definition of P that $P(x) \leqq \| x \|$, and therefore $F(x) \leqq p(x) \leqq k \| x \|$. Replacing x by $-x$ we get $F(x) \geqq -k \| x \|$ which together with $F(x) \leqq k \| x \|$ shows that $| F(x) | \leqq k \| x \|$, that is $\| F \| \leqq k$.

It is not without interest to notice that, in the ordered spaces case, a linear positive continuous functional does not need to possess a linear positive continuous extension without having to increase its norm. For instance, if S is the plane with its usual ordering and norm, and V is the set of all $x = (\xi, -\xi)$, then the functional f defined over V by $f(x) = \xi$ is positive (because the ordering on V is discrete) and has norm $\| f \| = 1/\sqrt{2}$, but any positive extension F of f has norm $\| F \| \geqq 1$.

We shall now prove Theorem 1.

Let E be a compact space. If $x \in c(E)$, then $\| x \| = \| x_+ \| \vee \| x_- \|$, where $x_+ = x \vee 0$ and $x_- = x \wedge 0$. Let us assume that $\| x_+ \| \geqq \| x_- \|$ so that $\| x \| = \| x_+ \|$. If $y \in c(E)$ and $y \geqq x$, then $y_+ \geqq x_+ \geqq 0$ and therefore $\| y_+ \| \geqq \| x_+ \|$ from which it follows $\| y \| \geqq \| y_+ \| \geqq \| x_+ \| = \| x \|$ and x is semi-positive. In the same way we see that, if $\| x_- \| \geqq \| x_+ \|$ then x is semi-negative. From this it follows immediately that condition (1) is necessary. The necessity of condition (2) is clear.

---

[6] See Banach [2], p. 55.

[7] A linear functional F is called positive if $F(x) \geqq 0$ whenever $x \geqq 0$.

[8] See Banach [2], p. 28.

Now let us assume that S satisfies conditions (1) and (2). Let S^* be the dual of S, that is the vector space of all linear continuous functionals over S endowed with its natural norm and ordering. We shall also consider on S^* its weak topology obtained by making the points of S act on S^* as linear functionals: it is known that the unit sphere U of S^* is weakly compact.[9] Moreover the set P of all positive elements of S^* is clearly weakly closed. Putting $E = U \cap P$ we get a compact space.

For every $x \varepsilon S$ let X be the real-valued function defined over E by $X(f) = f(x)$, where $f \varepsilon E$. It is clear that X is continuous over E and that the mapping $x \rightarrow X$ is a linear transformation from S into $c(E)$.

The mapping $x \rightarrow X$ is isometric. In fact, we have $| X(f) | = | f(x) | \leqq \| f \| \cdot \| x \| \leqq \| x \|$ and therefore $\| X \| \leqq \| x \|$. On the other hand, every $x \varepsilon S$ is either semi-positive or semi-negative. Assume that we are in the first case and $x \neq 0$ (the case $x = 0$ being trivial). Let V be the set of all $v = \lambda x$ and let f be defined over V by $f(v) = \lambda \| x \|$. If $v \varepsilon V$, $w \varepsilon S$, $v \leqq w$ then $f(v) \leqq \| w \|$. In fact, we have $v = \lambda x$: if $\lambda \leqq 0$ then $f(v) = \lambda \| x \| \leqq \| w \|$ is clear, and if $\lambda > 0$ then $v = \lambda x \leqq w$, that is $x \leqq w/\lambda$ implies $\| x \| \leqq \| w/\lambda \|$ because x is semi-positive and $f(v) \leqq \| w \|$ is still true. This being so, we may apply Theorem 2 and get a linear positive extension F of f defined over S such that $\| F \| \leqq 1$. Then $F \varepsilon E$ and $X(F) = F(x) = f(x) = \| x \|$. If x, instead of being semi-positive, is semi-negative we may get some $F \varepsilon E$ with $X(F) = - \| x \|$. In any case we see that $\| X \| \geqq \| x \|$. The final conclusion is that $\| X \| = \| x \|$.

The mapping $x \rightarrow X$ is one-to-one between S and its range: this follows from the above result.

Finally the mapping $x \rightarrow X$ is order-preserving in both directions. If $x \geqq 0$ then $X(f) = f(x) \geqq 0$, that is $X \geqq 0$. Assume now that $X \geqq 0$. We want to show that $x \geqq 0$ and therefore may assume $x \neq 0$. Put $\delta = P(-x)$, where P is the function introduced in the proof of Theorem 2. Let V be the set of all $v = \lambda x$ and let f be defined over V by $f(v) = - \lambda \delta$. If $v \varepsilon V$, $w \varepsilon S$, $v \leqq w$ then $f(v) \leqq \| w \|$. In fact, if $v = \lambda x$ and $\lambda \geqq 0$ then $f(v) = - \lambda \delta \leqq \| w \|$ is clear, and if $\lambda < 0$ then $v = \lambda x \leqq w$ implies $-x \leqq -w/\lambda$ and therefore $P(-x) \leqq \| w/\lambda \|$, by the definition of $P(-x)$, and $f(v) \leqq \| w \|$ still is true. Applying once more Theorem 2 we obtain a linear positive extension F of f defined over S with $\| F \| \leqq 1$. But then $F \varepsilon E$ and $X(F) = f(x) = - \delta$ and we must have $\delta = 0$, that is, $P(-x) = 0$. This implies the existence of a sequence $y_n \varepsilon S$ $(n = 1, 2, \cdots)$ such that

---

[9] See Bourbaki [4] and also Alaoglu [1].

$x+y_n \geqq 0$ and $x+y_n \to x$. By condition (2) we conclude that $x \geqq 0$. Theorem 1 is therefore proved.

An alternative form for Theorem 1 is the following:

THEOREM 1′. *In order that S be isomorphic in the order, vector and norm sense to a normed vector ordered sub-space of $c(E)$, for some compact space E, it is necessary and sufficient that:* (1′) *if $x \leqq y \leqq z$ then $\| y \| \leqq \| x \| \vee \| z \|$, and* (2′) *the set of all positive elements of S be closed.*

It is sufficient to notice that conditions (1) and (1′) are equivalent.

The idea of using the notions of semi-positive and semi-negative elements is also suitable for the lattice case. Let us assume that S is a normed (real) vector space and at the same time a vector lattice.[10] Then we have:

THEOREM 3. *In order that S be isomorphic in the lattice, vector and norm sense to a normed vector sub-lattice of $c(E)$, for some compact space E, it is necessary and sufficient that:* (1) *every element of S be either semi-positive or semi-negative, and* (2) $\| x \| \geqq \| \, | x | \, \|$, *where* $| x | = x \vee -x$.

The necessity of condition (1) was already proved, and that of condition (2) is clear because it is even true that $\| \, | x | \, \| = \| x \|$.

Now assume that (1) and (2) hold good. If $0 \leqq x \leqq y$ then $\| x \| \leqq \| y \|$. In fact, this is clear if x is semi-positive, and if x is semi-negative then $0 \leqq x$ implies $\| x \| \leqq 0$, that is, $x = 0$ and the result is still true.

Since $0 \leqq x_+ \leqq | x |$, $0 \leqq -x_- \leqq | x |$, we see that $\| x_+ \| \vee \| x_- \| \leqq \| \, | x | \, \|$.

If x is semi-positive, from $x \leqq x_+$ we get $\| x \| \leqq \| x_+ \|$. If x is semi-negative, the relation $x \geqq x_-$ implies $\| x \| \leqq \| x_- \|$. In any case we have $\| x \| \leqq \| x_+ \| \vee \| x_- \|$.

Combining these last two results with condition (2) we see that $\| x \| = \| \, | x | \, \|$ and that $\| x \| = \| x_+ \| \vee \| x_- \|$.

The fact that $0 \leqq x \leqq y$ implies $\| x \| \leqq \| y \|$, and that $\| x \| = \| \, | x | \, \|$ means that S is a *normed lattice,* which in addition has the property $\| x \| = \| x_+ \| \vee \| x_- \|$. Incidentally we notice that, conversely, every normed lattice with this additional property satisfies conditions (1) and (2) of Theorem 3.

Now we observe that, by a result of Bohnenblust and Kakutani, it is sufficient to prove that $x \wedge y = 0$ implies $\| x \vee y \| = \| x \| \vee \| y \|$. Since the condition $x \wedge y = 0$ is necessary and sufficient for the existence of a $z \varepsilon S$

---

[10] That is, in addition to the properties stated in [4], S is a lattice (see Birkhoff [3]).

with $z_+ = x$, $z_- = -y$, in which case $z = x - y$ and $|z| = x + y = x \vee y$, we see that the Bohnenblust-Kakutani condition is equivalent to $\| \, |z| \, \| = \| z_+ \| \vee \| z_- \|$, that is, to $\| z \| = \| z_+ \| \vee \| z_- \|$. Theorem 3 is therefore proved.

An alternative form for Theorem 3 is obtained by replacing condition (1) by its equivalent (1′).

UNIVERSITY OF CHICAGO.

BIBLIOGRAPHY.

1. L. Alaoglu, "Weak topologies of normed linear spaces," *Annals of Mathematics*, vol. 41 (1940), pp. 252-267.

2. S. Banach, *Théorie des opérations linéaires*, Warszawa, 1932.

3. G. Birkhoff, *Lattice Theory*, New York, 1940.

4. N. Bourbaki, "Sur les espaces de Banach," *Comptes Rendus de l'Académie des Sciences de Paris*, vol. 206 (1938), pp. 1701-1704.

5. H. F. Bohnenblust and S. Kakutani, "Concrete representations of (*M*)-spaces," *Annals of Mathematics*, vol. 42 (1941), pp. 1025-1028.

6. S. Kakutani, "Weak topology, bicompact sets and the principle of duality," *Proceedings of the Imperial Academy of Tokyo*, vol. 16 (1940), pp. 63-67.

7. S. Kakutani, "Concrete representations of abstract (*M*)-spaces," *Annals of Mathematics*, vol. 42 (1941), pp. 994-1024.

8. M. Krein, "Propriétés fondamentales des ensembles coniques normaux dans l'espace de Banach," *Comptes Rendus de l'Académie des Sciences de l'URSS*, vol. 28 (1940), pp. 13-17.

9. M. Krein and S. Krein, "On an inner characteristic of the set of all continuous functions defined on a bicompact Hausdorff space," *Comptes Rendus de l'Académie des Sciences de l'URSS*, vol. 27 (1940), pp. 427-430.

10. M. Krein et S. Krein, "Sur l'espace des fonctions continues définies sur un bicompact de Hausdorff et ses sousespaces semiordonnés," *Recueil Mathématique*, vol. 13 (1943), pp. 1-37.

FREE SUMS OF GROUPS AND THEIR GENERALIZATIONS. AN ANALYSIS OF THE ASSOCIATIVE LAW.*

By Reinhold Baer.

Introduction. Our problem and our method of approach can best be described in a situation in which the problems have been solved for a long time: the free sum of two groups with an amalgamated subgroup. (Note that the composition of group elements will be called addition in spite of the fact that this addition need not be commutative.) In this case there are given two groups G and H and a pair of isomorphic subgroups G' and H' of G and H respectively. If α is some fixed isomorphism of G' upon H', then we may identify the element x in G' and its image $x\alpha$ in H'; and we are led to the following structure: a group G, a group H and a group S which is a common subgroup of G and H and which is exactly the intersection of the two groups G and H. Now we form the join J of the two sets G and H subject to the requirement $G \cap H = S$; and in J we may define an addition in the following obvious fashion: if x and y both belong to G (or both belong to H), then their sum $x + y$ is defined as in G (as in H); and their sum is undefined otherwise. It is clear that the sum of any two elements in J is either undefined or uniquely determined. The problem is to imbed this structure into as general a group as possible, a problem that has been solved by O. Schreier [1] by his construction of the free sum of G and H with the amalgamated subgroup S.

We may regain this problem in a somewhat more general form by the following observation: Let T be a non-vacuous subset of the group L; if x, y and $x + y$ belong to T, then the sum of the elements x and y in T is the uniquely determined element $x + y$; and if x, y belong to T, though $x + y$ does not belong to T, then we let the sum of x and y be undefined. Thus the subset T has been made into an additive structure, and one may ask how to characterize by inner properties those structures that may be imbedded into groups. As may be expected, such a general discussion will provide the proper framework within which to discuss problems of the type indicated in the first paragraph.

In the first part of the present investigation (sections **1** to **5**) we develop

---

* Received May 2, 1948.

a general method for attacking such problems. Here we shall be concerned mainly with the various possibilities for formulating the associative law of addition, if we assume that the sum of two elements may or may not be defined, but if defined is unique; and we shall investigate the effect of imposing these various laws on the problem of imbedding the given system into an additive manifold (which possesses an addition which may always be effected in one and only one fashion and which is associative in the customary sense). It turns out that such an imbedding may sometimes be effected and sometimes it is impossible; and if it is possible to effect this imbedding, then we may search for what one may call a "most general" containing manifold. If the term "most general" is identified with "freeness" as recently formulated by Bates [1], then imbeddability implies imbeddability into a most general manifold; but if we identify the term "most general" with Schreier's concept of "freeness," then it is easy to construct sums of groups which are "Bates free" but not "Schreier free."

In the second part we apply the results of the first five sections to the problem of constructing sums of groups. Here we obtain proofs of the known results (the theorems of O. Schreier [1] and H. Neumann [1]) and a variety of more general criteria. It seems to the author as if the results obtained by no means exhaust the possibilities inherent in the methods which have been developed here.

One word concerning our methods. In most proofs for the existence of sums of groups one attempts more or less forcefully to represent the elements of the sum by some sort of a normal form which has the disadvantage that the formal sum of two normal forms is not a normal form. Artin [2] gave recently a proof for the existence of the free sum of groups (without amalgamation) in which the elements in the sum are classes of similar "vectors." Since in this case the elements are "almost in normal form" anyway, the difference between these methods is not too great in this special case. But in our more general situation it seems to be much more advantageous to forget about normal forms altogether; and this we do.

1. Summation.

The general frame work for our discussion is provided by the following concept.

DEFINITION 1. *An add is a non vacuous system A of elements with one composition, called addition $a + b$, meeting the following requirement:*

(U) *If a and b are elements in A, then there exists at most one element c in A such that $a + b = c$.*

Thus the sum of the elements a and b in A is either undefined or else it is uniquely determined.

If $a_1, \cdots, a_n$ are n (equal or different) elements in the add A, then we define the *summation* $\mathfrak{S}(a_1, \cdots, a_n)$ as a subset of A in the following way by complete induction.

DEFINITION 2. (1) *If u is an element in A, then $\mathfrak{S}(u)$ consists of u.* (n) *Suppose that $1 < n$ and that $u_1, \cdots, u_n$ are in A. Then the element u in A belongs to $\mathfrak{S}(u_1, \cdots, u_n)$ if, and only if, there exists an integer i with $0 < i < n$ and elements u' and u'' in $\mathfrak{S}(u_1, \cdots, u_i)$ and $\mathfrak{S}(u_{i+1}, \cdots, u_n)$ respectively such that $u = u' + u''$.*

Sometimes we shall say $\mathfrak{S}_A$ instead of $\mathfrak{S}$ in order to indicate that the summation has to be effected within the add A and not within some more comprehensive add.

We note that $\mathfrak{S}(a)$ is always a one-element set, that $\mathfrak{S}(a, b)$ is either vacuous or the one-element set consisting of $a + b$ (if this sum is defined); and that $\mathfrak{S}(a, b, c)$ may consist of 0, 1 or 2 elements. More generally we have:

$$\mathfrak{S}_A(u_1, \cdots, u_n) \textit{ is a finite subset of } A.$$

This important fact is easily verified by complete induction.

LEMMA 1. *Suppose that $s_1, \cdots, s_n$ are elements in A, $1 < n$.*

(a) *If a belongs to $\mathfrak{S}(s_1, \cdots, s_n)$, then there exists an integer i with the properties: $0 < i < n$, $s_i + s_{i+1}$ exists in A, and a belongs to*

$$\mathfrak{S}(s_1, \cdots, s_{i-1}, s_i + s_{i+1}, s_{i+2}, \cdots, s_n).$$

(b) *If $0 < j < n$, and if $s_j + s_{j+1}$ exists in A, then*

$$\mathfrak{S}(s_1, \cdots, s_{j-1}, s_j + s_{j+1}, s_{j+2}, \cdots, s_n) \leq \mathfrak{S}(s_1, \cdots, s_n).$$

These two facts are easily verified by complete induction.

If the subset S of A is not vacuous, then S may be considered as *a subadd of A* where we define addition in S in the obvious fashion. It is, however, quite possible that the sum of the elements s' and s'' in S exists in A, but not in S (namely in case $s' + s''$ does not belong to S). This leads us to the following concept.

DEFINITION 3. *If S is a subset of A such that $s' + s''$ belongs to S whenever s' and s'' are in S and $s' + s''$ exists in A, then S is closed in A.*

Note that this closure property is weaker than the subgroup property (the negative integers, for example, form a closed subset of the group of integers).

Lemma 2. *If $s_1, \cdots, s_n$ are elements in the subadd S of the add A, then*

(a) $\mathfrak{S}_S(s_1, \cdots, s_n) \leq \mathfrak{S}_A(s_1, \cdots, s_n) \cap S$; *and*

(b) $\mathfrak{S}_S(s_1, \cdots, s_n) = \mathfrak{S}_A(s_1, \cdots, s_n) \cap S$ *whenever S is closed in A.*

The inequality (a) is an almost immediate consequence of Definition 2; and the equation (b) may be verified by an easy induction argument, based on Lemma 1.

2. The derived additive manifold. The following notations will prove very convenient in the future and will be used throughout. If Q is a set of elements, and if $v_1, \cdots, v_n$ $(0 < n)$ are elements in Q, then $v = (v_1, \cdots, v_n)$ *is a vector of length n over Q.* The element v_i in Q is the i-th coordinate of the vector v. The system of all vectors over Q will be denoted by $V(Q)$. In $V(Q)$ we define *addition* by the rule

(A) $\quad (v_1, \cdots, v_m) + (w_1, \cdots, w_n) = (v_1, \cdots, v_m, w_1, \cdots, w_n)$.

Addition in $V(Q)$ is unique, associative and the cancellation law holds too, so that $V(Q)$ is a semi-group. (Note that our "vectors" take the place of what has often been termed "words.")

If in particular there is given an add A, then we introduce a similarity between the vectors over A as follows:

(S. 1) *If a, b, c are elements in A such that $a + b = c$, and if u and v are vectors over A, then the vectors $u + (a, b) + v$ and $u + (c) + v$ are directly similar.* (Here, as in similar circumstances, the vectors u and v may be of "length 0.")

This direct similarity should be interpreted symmetrically in the sense that direct similarity of the vectors r and s is equivalent to direct similarity of the vectors s and r. Direct similarity is not transitive; and thus we complement the rule (S. 1) by the following rule.

(S. 2) *The vectors v and w over A are similar, in symbols: $v \sim w$, if there exists a finite number of vectors $v(1), \cdots, v(k)$ over A, $0 < k$ such that $v = v(1)$, $v(i)$ and $v(i + 1)$ are directly similar for $0 < i < k$ and $v(k) = w$.*

It is clear that similarity of vectors is reflexive, symmetric, transitive. Thus we may form classes of similar vectors in $V(A)$; and we shall denote the class of similar vectors which contains the vector $v = (v_1, \cdots, v_n)$ by $\langle v \rangle = \langle (v_1, \cdots, v_n) \rangle$.

It is furthermore quite easy to see that sums of similar vectors are similar; in symbols: $v \sim v'$ and $w \sim w'$ imply $v + w \sim v' + w'$. Consequently it is possible to define addition of classes of similar vectors by means of the formula:

(A. 1) $\langle v \rangle + \langle w \rangle = \langle v + w \rangle$.

Denoting by $D(A)$, for A any add, the system of classes of similar vectors, subject to the definition A. 1 of addition, one verifies readily the following facts:

Addition of any two elements in $D(A)$ is well defined and addition in $D(A)$ is associative.

Thus we may call $D(A)$ *the derived additive manifold* (of the add A) in accordance with the following

DEFINITION 1. *An additive manifold is a nonvacuous system M of elements with one composition, called addition $a + b$, meeting the following requirements:*

(a) *If a and b are elements in M, then there exists one and only one element c in M such that $a + b = c$.*

(b) $a + (b + c) = (a + b) + c$ *for a, b, c in M.*

Because of the associative law (b) we may omit parentheses in sums of elements in M. It should be noted that we do not require validity of the cancellation law or commutative law.

To enunciate the fundamental properties of the derived manifold we need the concept of homomorphism.

DEFINITION 2. *If A and B are adds, if $\mathfrak{h}$ is a single valued A to B mapping, and if*

$$a + b = c \text{ for } a, b, c \text{ in } A \text{ implies } a\mathfrak{h} + b\mathfrak{h} = c\mathfrak{h},$$

then $\mathfrak{h}$ is a homomorphism of A into B.

If, for instance, the sum of elements in A never exists, then every single valued mapping of A into some add is a homomorphism.

THEOREM 1. *If A is an add, then the mapping*

(N) $a^* = \langle (a) \rangle$ *for a in A*

has the following properties:

(a) *Mapping x in A upon x^* in $D(A)$ is a homomorphism of the add A into the derived additive manifold $D(A)$.*

(b) *Every element in $D(A)$ is the sum of elements in the subset A^* (of all the elements x^* for x in A).*

(c) *If $\mathfrak{h}$ is a homomorphism of A into the additive manifold M, then there exists one and only one homomorphism $\mathfrak{k}$ of $D(A)$ into M such that $x^*\mathfrak{k} = x\mathfrak{h}$ for every x in A.*

Proof. It is clear that a single valued mapping of A into $D(A)$ is effected by mapping x in A onto $x^* = \langle (x) \rangle$ in $D(A)$. If a, b, c are elements in A such that $a + b = c$, then it follows from the preceding definitions that

$$a^* + b^* = \langle (a) \rangle + \langle (b) \rangle = \langle (a) + (b) \rangle = \langle (a, b) \rangle = \langle (c) \rangle = c^*.$$

This shows the validity of (a).

If y is an element in $D(A)$, then there exist elements $a_1, \cdots, a_n$ in A such that

$$\begin{aligned} y &= \langle (a_1, \cdots, a_n) \rangle = \langle (a_1) + \cdots + (a_n) \rangle \\ &= \langle (a_1) \rangle + \cdots + \langle (a_n) \rangle = a^*_1 + \cdots + a^*_n, \end{aligned}$$

and this proves the validity of (b).

If $\mathfrak{h}$ is a homomorphism of A into the additive manifold M, and if $\mathfrak{k}$ is a homomorphism of $D(A)$ into M such that $x^*\mathfrak{k} = x\mathfrak{h}$ for x in A, then we have, for $a_1, \cdots, a_n$ in A, as before

$$\begin{aligned} \langle (a_1, \cdots, a_n) \rangle \mathfrak{k} &= (a^*_1 + \cdots + a^*_n)\mathfrak{k} = a^*_1\mathfrak{k} + \cdots + a^*_n\mathfrak{k} \\ &= a_1\mathfrak{h} + \cdots + a_n\mathfrak{h}. \end{aligned}$$

Thus $\mathfrak{k}$ is uniquely determined by $\mathfrak{h}$, if it exists at all. To prove the existence of at least one homomorphism $\mathfrak{k}$, meeting our requirements, we proceed as follows:

If $a_1, \cdots, a_n$ are elements in A, then $a_1\mathfrak{h} + \cdots + a_n\mathfrak{h}$ is a well determined element in the additive manifold M. Thus we may let

$$(a_1, \cdots, a_n)\mathfrak{h} = a_1\mathfrak{h} + \cdots + a_n\mathfrak{h}.$$

If a, b, c are elements in A such that $a + b = c$, then $a\mathfrak{h} + b\mathfrak{h} = c\mathfrak{h}$ since $\mathfrak{h}$ is a homomorphism. If u and v are vectors over A, then it follows now from the validity of the associative law in M that

$$\begin{aligned} [u + (a, b) + v]\mathfrak{h} &= u\mathfrak{h} + (a, b)\mathfrak{h} + v\mathfrak{h} = u\mathfrak{h} + a\mathfrak{h} + b\mathfrak{h} + v\mathfrak{h} \\ &= u\mathfrak{h} + c\mathfrak{h} + v\mathfrak{h} = [u + (c) + v]\mathfrak{h}. \end{aligned}$$

Hence $\mathfrak{h}$ maps directly similar vectors upon the same element. Consequently any two similar vectors are mapped by $\mathfrak{h}$ upon the same element in M. Hence a single valued mapping $\mathfrak{k}$ of $D(A)$ into M is defined by the rule:

$$\langle v\rangle\mathfrak{k} = v\mathfrak{h} \text{ for } v \text{ in } V(A).$$

From $\langle v\rangle\mathfrak{k} + \langle w\rangle\mathfrak{k} = v\mathfrak{h} + w\mathfrak{h} = (v+w)\mathfrak{h} = \langle(v+w)\rangle\mathfrak{k}$, by the associative law in M, we deduce finally that $\mathfrak{k}$ is a homomorphism of $D(A)$ into M. If x is an element in A, then we have $x^*\mathfrak{k} = \langle(x)\rangle\mathfrak{k} = (x)\mathfrak{h} = x\mathfrak{h}$; and this completes the proof.

The homomorphism of A into $D(A)$ which is defined by Theorem 1, (N) will be referred to as *the natural homomorphism of the add A into its derived additive manifold $D(A)$.*

In order to show that the derived manifold and the natural homomorphism are essentially uniquely determined by the properties (b) and (c) of Theorem 1, we need the concept of isomorphism.

DEFINITION 3. *An isomorphism of the add A upon the add B is a correspondence $\mathfrak{a}$ meeting the following requirements:*

(a) *$\mathfrak{a}$ is a one to one mapping of A upon $B = A\mathfrak{a}$;*

(b) *if a, b, c are elements in A, then the validity of $a + b = c$ is necessary and sufficient for the validity of $a\mathfrak{a} + b\mathfrak{a} = c\mathfrak{a}$.*

Thus if $\mathfrak{a}$ is an isomorphism of A upon B, then there exists the inverse correspondence $\mathfrak{a}^{-1}$ and is an isomorphism of B upon A. It should in particular be noted that a homomorphism of A into B which is at the same time a one-to-one mapping of A upon B need not be an isomorphism. If, for instance, the adds A and B consist of the same elements, though using different definitions of additions, then it is quite possible that the identity mapping is a homomorphism of A upon B without being an isomorphism of A upon B.

THEOREM 2. *Suppose that the homomorphism $\mathfrak{r}$ of the add A into the additive manifold M has the following properties.*

(a) *Every element in M is the sum of elements in $A\mathfrak{r}$.*

(b) *If $\mathfrak{h}$ is a homomorphism of A into some additive manifold N, then there exists a homomorphism $\mathfrak{k}$ of M into N such that $\mathfrak{h} = \mathfrak{r}\mathfrak{k}$.*

Then there exists an isomorphism $\mathfrak{v}$ of M upon $D(A)$ such that $x^* = x\mathfrak{r}\mathfrak{v}$ for x in A.

Proof. We apply first condition (b) upon the natural homomorphism of A into $D(A)$. Hence there exists a homomorphism $\mathfrak{d}$ of M into $D(A)$ such that $x^* = x\mathfrak{r}\mathfrak{d}$ for x in A. Next we apply Theorem 1, (c) upon the homomorphism $\mathfrak{r}$ of A into M. Thus there exists a homomorphism $\mathfrak{k}$ of $D(A)$ into M such that $x^*\mathfrak{k} = x\mathfrak{r}$ for x in A. We have consequently

$$x^* = x\mathfrak{r}\mathfrak{d} = x^*\mathfrak{k}\mathfrak{d} \text{ and } x\mathfrak{r} = x^*\mathfrak{k} = x\mathfrak{r}\mathfrak{d}\mathfrak{k} \text{ for every } x \text{ in } A.$$

If y is an element in $D(A)$, then it follows from Theorem 1, (b) that y is the sum of elements in A^*. But every element x^* in A^* is left invariant by $\mathfrak{k}\mathfrak{d}$; and thus it follows that $y = y\mathfrak{k}\mathfrak{d}$ for y in $D(A)$. Likewise we deduce from our hypothesis (a) that $z = z\mathfrak{d}\mathfrak{k}$ for every z in M. Thus we have shown that $\mathfrak{k}\mathfrak{d} = 1$ and $\mathfrak{d}\mathfrak{k} = 1$. But $\mathfrak{k}$ and $\mathfrak{d}$ are homorphisms and consequently they are reciprocal isomorphisms between $D(A)$ and M. Hence $\mathfrak{d}$ is the desired isomorphism of M upon $D(A)$.

3. The associative laws. We are now going to enunciate a variety of properties of adds which would be equivalent to the customary associative law, if the sum of any two elements in the add would always exist. But as properties of adds they can be shown not to be equivalent.

If v is a vector over the add A, then $v = (v_1, \cdots, v_n)$; and thus the addition $\mathfrak{S}v = \mathfrak{S}(v_1, \cdots, v_n)$ is well defined (see section **1**).

THE FIRST OR WEAK ASSOCIATIVE LAW *is satisfied by the add A, if $\mathfrak{S}v$ contains, for every vector v over A, at most one element.*

We note that $\mathfrak{S}v$ contains one and only one element, if v has length 1; and that $\mathfrak{S}v$ contains at most one element, if v is of length 2. For vectors of length 3 the weak associative law is rather similar in effect to the customary associative law.

THE SECOND ASSOCIATIVE LAW *is satisfied by the add A, if the natural homomorphism of A into its derived additive manifold $D(A)$ is a one-to-one correspondence.*

The following fact will be useful.

LEMMA 1. *If v is a vector over the add A, and if the element u in A belongs to $\mathfrak{S}v$, then $(u) \sim v$.*

This is easily verified by complete induction with respect to the length of v. (Use **1**, Lemma 1, (a).)

The first associative law is a consequence of the second associative law.

Proof. If the elements a and b belong to $\mathfrak{S}v$, then it follows from Lemma 1 that $(a) \sim v \sim (b)$. Hence $a^* = b^*$; and $a = b$ is a consequence of the fact that the natural homomorphism is one to one. Hence $\mathfrak{S}v$ contains at most one element.

In general, the second associative law is not a consequence of the first one. However, we shall show (in section **4**) that they are equivalent in the cases of greatest interest.

THE THIRD OR INTERMEDIATE ASSOCIATIVE LAW *is satisfied by the add A, if the natural homomorphism of A into its derived additive manifold is an isomorphism.*

It is obvious that the second associative law is a consequence of the third one: and their non-equivalence may be verified easily [see **6**, Example 2]. We may decompose the third associative law into two properties as follows:

(III. 1) $(a) \sim (b)$ *if, and only if,* $a = b$.

(III. 2) $(a, b) \sim (c)$ *if, and only if,* $a + b = c$.

(III. 1) is equivalent to the second associative law, since $(a) \sim (b)$ if, and only if, $a^* = b^*$. (III. 1) and (III. 2) together are equivalent with the third associative law, since $(a, b) \sim (c)$ if, and only if, $a^* + b^* = c^*$ (see also **2**, Definition 3).

THE FOURTH OR STRONG ASSOCIATIVE LAW *is satisfied by the add A, if* $v \sim w$ *implies always* $\mathfrak{S}v = \mathfrak{S}w$.

The third associative law is a consequence of the fourth associative law.

Proof. If $(a) \sim (b)$, then we infer $\mathfrak{S}(a) = \mathfrak{S}(b)$ from the strong associative law. This implies $a = b$, since $\mathfrak{S}(x)$ consists of the element x alone. Hence (III. 1) is true. If $(a, b) \sim (c)$, then we infer $\mathfrak{S}(a, b) = \mathfrak{S}(c)$ from the strong associative law. Since $\mathfrak{S}(c)$ consists of c only, it follows that c belongs to $\mathfrak{S}(a, b)$; and this implies $a + b = c$. Hence (III. 2) is true too.

The converse of the preceding statement is not true [**6**, Example 3]; and it appears that there exists a considerable gap between the third and the fourth associative law.

In what follows we shall refer to the various associative laws either by their names or by roman numerals. We remark that I seems to be the most natural generalization of the usual form of the associative law, as usually formulated. The importance of III is fairly apparent; and, as a matter of fact, IV will be by far the most important of all. The importance of III and IV will be elucidated somewhat in section **5**. Some indication of the importance of Properties (II) and (III) is contained in the following useful proposition which is a generalization of a result due to H. Neumann ([1], p. 598, Theorem).

LEMMA 2. *The natural homomorphism of the add A into $D(A)$ is an isomorphism (is one to one) if there exists an isomorphism (a one-to-one homomorphism) of A into some additive manifold.*

Proof. If $\mathfrak{h}$ is a homomorphism of A into the additive manifold M, then we deduce from **2**, Theorem 1, (c) the existence of a homomorphism $\mathfrak{k}$ of $D(A)$ into M such that $x^*\mathfrak{k} = x\mathfrak{h}$ for x in A. Now it is clear that the natural homomorphism is one to one, if $\mathfrak{h}$ is one to one; and the natural homomorphism is an isomorphism whenever $\mathfrak{h}$ is an isomorphism.

COROLLARY. *The add A may be imbedded into an additive manifold if, and only if, the third associative law is satisfied by A.*

This is a fairly immediate consequence of Lemma 2 and the existence of the derived manifold $D(A)$.

PROPOSITION 1. (a) *If the first (or second or third) associative law is satisfied by the add A, then it is also satisfied by the subadds of A.*
(b) *If the strong associative law is satisfied by the add A, then it is also satisfied by the closed subadds of A.*

Proof. If the weak associative law is satisfied by A, if S is a subadd of A, and if v is a vector over S, then $\mathfrak{S}_S v \leq \mathfrak{S}_A v$ [**1**, Lemma 2, (a)]. Since the latter summation contains at most one element, so does the former. Hence the weak associative law is satisfied by S.

If the natural homomorphism of A into $D(A)$ is one to one (is an isomorphism), then it induces a one-to-one homomorphism (an isomorphism) of the subadd S of A into the additive manifold $M(A)$. It follows from Lemma 2 that the natural homomorphism of S is one to one (is an isomorphism); and this completes the proof of (a).

Assume finally the validity of the strong associative law in A; and suppose that the subadd S of A is closed in A. If v and w are similar vectors over S, then it follows from $V(S) \leq V(A)$ that v and w are similar vectors over A. Hence $\mathfrak{S}_A v = \mathfrak{S}_A w$, since the strong associative law holds in A. Since S is closed in A, we deduce

$$\mathfrak{S}_S v = \mathfrak{S}_A v \cap S = \mathfrak{S}_A w \cap S = \mathfrak{S}_S w$$

from 1, Lemma 2, (b); and this shows the validity of the strong associative law in S.

Remark 1. It is not difficult to construct subadds of groups which do not satisfy the strong associative law (see, for instance, **6**, Example 3). This shows that Proposition 1, (b) would fail to be true, if we omitted the requirement that S be closed in A.

PROPOSITION 2. *The associative law* (X), *for* X = I, II, III, IV, *is satisfied by the add A, if every finite subset of A is contained in a subadd of A with Property* (X).

Proof. Assume that every finite subset of A is contained in a subadd of A which satisfies the strong associative law. Suppose that $v = (v_1, \cdots, v_n)$ and that the vector w of length $n - 1$ is directly similar to v. Form a finite subset F of A as follows: F contains every v_i; it contains $v_i + v_{i+1}$, if it exists in A; it contains $(v_i + v_{i+1}) + v_{i+2}$ and $v_{j-1} + (v_j + v_{j+1})$, if these sums exist in A; and so on. It is clear that F is finite, that v and w are vectors over F, and that $\mathfrak{S}_A v = \mathfrak{S}_F v$ and $\mathfrak{S}_A w = \mathfrak{S}_F w$. But F is, by hypothesis, contained in a subadd S of A in which the strong associative law is satisfied. Now one sees that $\mathfrak{S}_A v = \mathfrak{S}_F v = \mathfrak{S}_S v = \mathfrak{S}_S w = \mathfrak{S}_F w = \mathfrak{S}_A w$. Hence directly similar vectors in A have the same summation in A; and consequently any two similar vectors in A have the same summation in A. Thus the strong associative law is valid in A. The other contentions of our proposition are verified in a similar fashion.

DEFINITION. *The vector $v = (v_1, \cdots, v_n)$ over A is contractable, if $1 < n$ and if there exists an integer i such that $0 < i < n$ and such that the sum $v_i + v_{i+1}$ exists in A.*

Using the terminology of section **2** we may also say that the vector v of length n is contractable, if $1 < n$ and if there exists a vector w of length $n - 1$ which is directly similar to v.

THEOREM 1. *The following properties of the add A imply each other.*

(i) *If a is in A and v in $V(A)$, and if $(a) \sim v$, then a belongs to $\mathfrak{S}v$.*

(ii) *If $a, a_1, \cdots, a_n$ are elements in A such that $(a) \sim (a_1, \cdots, a_n)$, then $n = 1$ and $a = a_1$ or else $1 < n$ and the vector $(a_1, \cdots, a_n)$ is contractable.*

(iii) *If v and w are similar vectors over A, and if $\mathfrak{S}v$ is not vacuous, then $\mathfrak{S}w$ consists of one and only one element.*

(iv) *The weak associative law holds in A; and if v and w are directly similar vectors of length $n-1$ and n respectively such that $\mathfrak{S}w$ is not vacuous, then $\mathfrak{S}v$ is not vacuous.*

(v) *The strong associative law holds in A.*

Proof. Assume the validity of (i); and suppose that $a, a_1, \cdots, a_n$ are elements in A such that $(a) \sim (a_1, \cdots, a_n) = v$. It follows from (i) that a belongs to $\mathfrak{S}v$. If $n = 1$, then $\mathfrak{S}v$ consists of a_1 alone so that $a = a_1$; and if $1 < n$, then we infer the contractability of v from **1**, Lemma 1, (a). Thus (ii) is a consequence of (i).

Assume next the validity of (ii). Suppose that r is a vector over A and that a and b are elements in $\mathfrak{S}r$. It follows from Lemma 1 that $(a) \sim r \sim (b)$; and hence it follows from (ii) that $a = b$. This proves the validity of the weak associative law in A. Consider next a vector v over A such that $\mathfrak{S}v$ is not vacuous. If there existed among the vectors similar to v one with vacuous summation, then there would exist a vector w of minimal length m such that $v \sim w$ and $\mathfrak{S}w$ is vacuous. Clearly $1 < m$, since $\mathfrak{S}(x)$, for x in A, consists of x. There exists an element u in $\mathfrak{S}v$; and it follows from Lemma 1 that $(u) \sim v$. Consequently $(u) \sim w$; and we infer from (ii) the contractability of w. Hence there exists a vector s of length $m-1$ which is directly similar to w; and it follows from **1**, Lemma 1, (b) that $\mathfrak{S}s \leqq \mathfrak{S}w$. Since $v \sim w \sim s$, and since the length of s is less than the (minimal) length of w, it follows that $\mathfrak{S}s$ is not vacuous. Hence $\mathfrak{S}w$ cannot be vacuous either, a contradiction which proves that $\mathfrak{S}t$ is not vacuous, whenever $t \sim v$ (and $\mathfrak{S}v$ is not vacuous). Since we have already verified the validity of the weak associative law, we have shown that (iii) is a consequence of (ii).

(iv) is a special case of (iii) (and hence a consequence of (iii)).

Assume now the validity of (iv). Suppose that v and w are directly similar vectors of length $n-1$ and n respectively. Then it follows from **1**, Lemma 1, (b) that $\mathfrak{S}v \leq \mathfrak{S}w$. It follows from (iv) that either $\mathfrak{S}v$ and $\mathfrak{S}w$ are both vacuous or they contain both exactly one element. In either case we may infer $\mathfrak{S}v = \mathfrak{S}w$ from $\mathfrak{S}v \leq \mathfrak{S}w$. Thus the summations of directly similar vectors are equal; and it follows (by complete induction) that the summations of similar vectors are equal. Thus the strong associative law holds in A, proving that (v) is a consequence of (iv).

Assume finally the validity of the strong associative law. If a is in A and v in $V(A)$, and if $(a) \sim v$, then $\mathfrak{S}(a) = \mathfrak{S}v$. But $\mathfrak{S}(a)$ consists of a alone, showing that a belongs to $\mathfrak{S}v$. Thus (i) is a consequence of (v); and this completes the proof.

Remark 2. Condition (ii) may be restated in the following very useful form which involves the natural homomorphism of A into $D(A)$.

(ii′) *The natural homomorphism of A into $D(A)$ is one to one; and if $a_1, \cdots, a_n$ are elements in A such that $1 < n$ and such that the sums $a_i + a_{i+1}$, for $0 < i < n$, do not exist in A, then $a^*_1 + \cdots + a^*_n$ does not belong to A^*.*

The equivalence of the properties (ii) and (ii′) is easily deduced from the fact that

$$(a) \sim (a_1, \cdots, a_n) \text{ and } a^* = a^*_1 + \cdots + a^*_n$$

are equivalent properties of the $n + 1$ elements $a, a_1, \cdots, a_n$ in A.

4. Self-reflexive adds. This important class of adds is characterized by the validity of the following postulates.

(SR. 1) *There exists an element 0 in A such that $x = x + 0 = 0 + x$ for every x in A.*

It is clear that the null-element is uniquely determined, if it exists. It may happen that a subadd of an add with 0 may posess a null element, not 0. We shall therefore usually restrict our attention to those subadds of A which contain the element 0.

(SR. 2) *To every element x in A there exists an element y in A such that $x + y = 0$.*

(SR. 3) *$x + y = 0$ and $x + z = 0$ imply $y = z$.*

If (SR. 2) and (SR. 3) are satisfied, then we may denote the uniquely determined solution of the equation $a + x = 0$ by $x = -a$. Instead of $(-a) + b$ and $c + (-a)$ we shall write $-a + b$ and $c - a$.

(SR. 4) *$x + y = 0$ if, and only if, $y + x = 0$.*

Together with the preceding postulates this implies $-(-x) = x$.

If these four postulates are satisfied by the add A, then we shall term the add A *self-reflexive*. Sometimes it will be convenient to be able to use the following rule which will be satisfied in all important situations.

(SR. 5) *If $a + b = 0$, and if the sum $b + c$ exists in A, then $a + (b + c) = c$; if the sum $r + s$ exists in A, and if $s + t = 0$, then $(r + s) + t = r$.*

An add A which satisfies all the rules (SR. 1) to (SR. 5) shall be termed *self-reflexive in the strict sense.*

PROPOSITION 1. *If the add A has properties* (SR. 1) *and* (SR. 2), *then the derived additive manifold is a group.*

Proof. Addition in $D(A)$ is always defined and associative. It is clear that $\langle(0)\rangle$ is the null-element of the additive manifold $D(A)$. If d is an element in $D(A)$, then $d = \langle(d_1, \cdots, d_n)\rangle$ with d_i in A. Hence there exist elements d'_i in A such that $d_i + d'_i = 0$ for every i. Let $d' = \langle(d'_n, \cdots, d'_1)\rangle$. Then one verifies readily that

$$d + d' = \langle(d_1, \cdots, d_n, d'_n, \cdots, d'_1)\rangle = \langle(0)\rangle.$$

It is well known, and easily verified, that under these circumstances the additive manifold $D(A)$ is a group.

LEMMA 1. *Assume the validity of* (SR. 1) *and* (SR. 2) *in the add A.*

(a) (SR. 3) *is a consequence of the second associative law.*

(b) (SR. 4) *and* (SR. 5) *are consequences of the third associative law.*

Proof. If x, y, z are elements in A such that $x + y = x + z = 0$, then $x^* + y^* = x^* + z^* = 0^*$. But $D(A)$ is a group [Proposition 1]. Hence $y^* = z^*$ and $y = z$ is a consequence of (II).

If $x + y = 0$, then $x^* + y^* = 0^*$. Since $D(A)$ is a group, we have $y^* + x^* = 0^*$; and $y + x = 0$ is a consequence of (III). Likewise we infer from $x + y = 0$ and the existence of the sum $y + z$ that

$$x^* + (y + z)^* = x^* + y^* + z^* = (x + y)^* + z^* = 0^* + z^* = z^*$$

and that therefore $x + (y + z) = z$ [here we applied (III) twice]; and the other half of (SR. 5) is verified likewise.

For a convenient enunciation of the next proposition we define inductively the concept of *i-fold contraction of a vector.* Every vector is its own 0-fold contraction. If w is an i-fold contraction of v, if $w = r + (a, b) + s$ [where the vectors r and s may have length 0], and if the sum $a + b$ exists in A, then $r + (a + b) + s$ is an $(i + 1)$-fold contraction of the vector v. Instead of "i-fold contraction" we shall often say only "contraction" [see **3**, Definition].

One deduces easily from **2**, (S. 1) and **2**, (S. 2) that *vectors are similar to all their contractions.* The following property is often useful.

LEMMA 2. *The element a in A belongs to $\mathfrak{S}v$ if, and only if, the vector (a) is a contraction of the vector v.*

This is easily verified by complete induction.

LEMMA 3. *If the vector* $v(1)+\cdots+v(n)$ *is an* i*-fold contraction of the vector* w, *then there exist vectors* $w(j)$ *with the following properties:* $v(j)$ *is an* $i(j)$*-fold contraction of* $w(j)$, $w=w(1)+\cdots+w(n)$ *and* $i=i(1)+\cdots+i(n)$.

The proof is effected by complete induction with respect to i after direct verification of the cases $i=0$ and $i=1$.

THEOREM 1. *If the add* A *is self-reflexive in the strict sense, and if the vectors* v *and* w *over* A *possess a common contraction, then there exists a vector* z *over* A *such that both* v *and* w *are contractions of* z.

We precede the proof of Theorem 1 by the proof of the following special case.

(1. i) *If* $u+(a+b)+v$ *is an* i*-fold contraction of the vector* w, *then there exists a vector* z *such that both* $u+(a,b)+v$ *and* w *are contractions of* z.

Proof by complete induction with respect to i.

If $i=0$, $u+(a+b)+v=w$ is a 1-fold contraction of $z=u+(a,b)+v$. This shows the validity of (1. 0). Assume now that $0<i$ and that (1. j) has been verified for $0\leq j<i$. Suppose that $u+(a+b)+v$ is an i-fold contraction of the vector w. It follows from the definition of i-fold contraction that there exists a vector t such that $u+(a+b)+v$ is a 1-fold contraction of t and t is an $(i-1)$-fold contraction of w. We distinguish two cases.

Case 1. $t=u+(c,d)+v$ with $a+b=c+d=e$.

We infer from Lemma 3 the existence of vectors p, q, r with the following properties: u is a contraction of p, (c,d) is a contraction of q, v is a contraction of r; and $w=p+q+r$. Since A is self-reflexive in the strict sense, there exists a well determined element $-a$ in A which satisfies $a-a=0$ and $-a+e=-a+(a+b)=b$. We let $z=p+(a,-a)+q+r$. Then w is a 2-fold contraction of z. Since u, (c,d), v are contractions of p, q and r respectively, it follows that $u+(a,-a)+(c,d)+v$ is a contraction of z. Since $c+d=e$ and $-a+e=b$, it follows that $u+(a,b)+v$ is a 2-fold contraction of the latter vector. Hence both w and $u+(a,b)+v$ are contractions of z.

Case 2. t has not the form discussed in Case 1.

Since $u+(a+b)+v$ is a 1-fold contraction of t, we may assume without loss in generality that

$$t = p + (c, d) + q + (a + b) + v \text{ with } u = p + (c + d) + q.$$

Since t is an $(i-1)$-fold contraction of w, we infer from $(1. i-1)$ the existence of a vector z over A such that both w and $p + (c, d) + q + (a, b) + v$ are contractions of z. Since $u + (a, b) + v$ is a 1-fold contraction of $p + (c, d) + q + (a, b) + v$, it is likewise a contraction of z.

This completes the inductive proof of (1. i).

For convenience of inductive argument we restate Theorem 1 as follows:

(i. j) *If the vector c over A is an i-fold contraction of v and a j-fold contraction of w, then there exists a vector z over A such that both v and w are contractions of z.*

Proof by complete induction with respect to i.

If $i = 0$, then $v = c$ is a j-fold contraction of $w = z$. This shows the validity of (0. j). If $i = 1$, then $v = r + (a, b) + s$ and $c = r + (a + b) + s$ and c is a j-fold contraction of w. Thus we obtain for $i = 1$ just the assertion (1. j) which we have verified before.

Assume now that $1 < i$ and that (k. j) has been verified for $0 \leq k < i$ (and every j). If c is an i-fold contraction of the vector v and a j-fold contraction of the vector w, then there exists a vector p such that c is a 1-fold contraction of p and p is an $(i-1)$-fold contraction of v. Since c is a 1-fold contraction of p and a j-fold contraction of w, we infer from (1. j) the existence of a vector q over A such that both w and p are contractions of q. Since p is an $(i-1)$-fold contraction of v and an m-fold contraction of q, we infer from $(i-1. m)$, (the inductive hypothesis), the existence of a vector z over A such that both v and q are contractions of z. Since w is a contraction of the contraction q of z, it follows that w is a contraction of z. Thus v and w are contractions of z, as we intended to prove. This completes the proof of (i. j) and of Theorem 1.

Remark 1. It is worth noting that only the properties (SR. 1), (SR. 2), (SR. 4) and the first half of (SR. 5) have been used in the proof of Theorem 1.

COROLLARY 1. *The vectors v and w over the strictly self-reflexive add A are similar if, and only if, there exists a vector z over A such that both v and w are contractions of z.*

Proof. If v and w are contractions of z, then $v \sim z \sim w$, as has been pointed out before; and $v \sim w$ is a consequence of the transitivity of similarity. If conversely v and w are similar vectors, then one deduces readily from the definition of similarity the existence of vectors $v(i)$ with the following properties:

$v = v(0)$, $v(i)$ is a contraction of $v(i+1)$ or $v(i+1)$ is a contraction of $v(i)$, $v(n) = w$.

We are going to construct by complete induction with respect to i vectors $z(i)$ such that v and $v(i)$ are both contractions of $z(i)$. For $i = 0$ we may clearly choose $z(0) = v$. Thus suppose that we have already constructed a vector $z(i)$ of which v and $v(i)$ are contractions. If $i = n$, then we have obtained the desired vector z. If $i < n$, then we distinguish two cases.

Case 1. $v(i+1)$ is a contraction of $v(i)$.

Then $v(i+1)$ is also a contraction of $z(i)$; and so we may let $z(i+1) = z(i)$.

Case 2. $v(i+1)$ is not a contraction of $v(i)$.

Then $v(i)$ is a contraction of both $z(i)$ and $v(i+1)$; and it follows from Theorem 1 that there exists a vector $z(i+1)$ over A such that both $z(i)$ and $v(i+1)$ are contractions of $z(i+1)$. Since v is a contraction of $z(i)$, v is likewise a contraction of $z(i+1)$.

This completes the induction. Hence there exists a vector $z = z(n)$ of which both v and $v(n) = w$ are contractions. This completes the proof.

COROLLARY 2. *The first and second associative law are equivalent properties of self-reflexive adds in the strict sense.*

Proof. It has been shown in section **3** that the first associative law is a consequence of the second one. Assume now that A is a self-reflexive add in the strict sense and that the first associative law is satisfied by A. Consider elements a, b in A such that $a^* = b^*$. Then $(a) \sim (b)$; and it follows from Corollary 1 that there exists a vector v over A of which the vectors (a) and (b) are contractions. We infer from Lemma 2 that the elements a and b belong both to $\mathfrak{S}v$; and $a = b$ is now a consequence of the first associative law. Thus $a = b$ is a consequence of $a^* = b^*$; and this shows the validity of the second associative law.

Remark 2. It may be of interest to obtain results similar to the preceding ones without the use of (SR. 5). In this respect we state without proof the following fact:

The second associative law is satisfied by the self-reflexive add A (in the weak sense) if, and only if, the first associative law is satisfied by A and $(0) \sim (a, b)$ implies $0 = a + b$.

THEOREM 2. *If A is a self-reflexive add (in the weak sense), then the following conditions are necessary and sufficient for the validity of the strong associative law in A.*

(a) *A is a self-reflexive add in the strict sense.*

(b) *The weak associative law is satisfied by A.*

(c) *If $(0) \sim v$, and if the length of v is greater than 2, then v is contractable.*

Proof. Assume first the validity of the strong associative law. Then we have shown in section **3** that all four associative laws hold in A. Thus (b) is true; and the validity of (a) is a consequence of Lemma 1. The validity of (c) may be inferred from **3**, Theorem 1.

Assume conversely the validity of the conditions (a), (b), (c). Then it follows from (a), (b) and Corollary 2 that the second associative law holds in A. If a and b are elements in A such that $(a) \sim (b)$, then $a^* = b^*$; and $a = b$ is a consequence of the second associative law. Next we prove by complete induction with respect to n (for $1 < n$) the validity of the following fact.

(C. n) If the vector v of length n is similar to a vector of length 1, then v is contractable.

To prove (C. 2) consider elements a, b, c in A such that $(a) \sim (b, c)$. Hence $(0) \sim (-a, a) \sim (-a, b, c)$; and it follows from (c) that at least one of the sums $-a + b$ and $b + c$ exists in A. If $-a + b$ exists in A, then $(-c) \sim (-a, b, c, -c) \sim (-a + b, 0) \sim (-a + b)$; and $-c = -a + b$ is a consequence of the second associative law. Applying (SR. 5) twice we find that $a - c = a + (-a + b) = b$ and $b + c = (a - c) + c = a$. If $-a + b$ does not exist in A, then $b + c$ exists in A. Hence $(a) \sim (a, -a, b, c) \sim (0, b + c) \sim (b + c)$; and this implies, as has been pointed out before, $a = b + c$. Thus we have in either case $a = b + c$. Consequently (C. 2) (and the third associative law) are true.

Assume now that $2 < n$ and that the validity of (C. j) has been verified for $2 \leq j < n$. Assume that $a, b_1, \cdots, b_n$ are elements in A such that $(a) \sim (b_1, \cdots, b_n)$. Then $(0) \sim (-a, a) \sim (-a, b_1, \cdots, b_n)$; and we infer from Condition (c) the existence of at least one of the sums $-a + b_1$, $b_1 + b_2, \cdots, b_{n-1} + b_n$ in A. If $-a + b_1$ exists in A, then

$$\begin{aligned}(-(-a + b_1)) &\sim (-(-a + b_1), -a, b_1, \cdots, b_n) \\ &\sim (-(-a + b_1), -a + b_1, b_2, \cdots, b_n) \sim (b_2, \cdots, b_n);\end{aligned}$$

and it follows from (C. n — 1) that $(b_2, \cdots, b_n)$ is contractable. Thus the vector $(b_1, \cdots, b_n)$ is in all cases contractable. This proves (C. n) and completes the induction.

Now we have verified the validity of Condition (ii) of **3**, Theorem 1; and this shows the validity of the strong associative law in A, as we claimed.

COROLLARY 3. *The strong associative law is satisfied by the strictly self-reflexive add if, and only if,* (0) *is a contraction of the vector* v *whenever* $(0) \sim v$.

Proof. The necessity of our condition is a consequence of the fact that $\mathfrak{S}(0)$ consists of 0 alone; and that (0) is a contraction of v whenever 0 belongs to $\mathfrak{S}v$. To prove the sufficiency of our condition we have to verify only the validity of the second associative law (by Theorem 2); and this is easily done, since $a^* = b^*$ implies $(a) \sim (b)$, and since this implies $(0) \sim (b, -a)$ so that (0) is a contraction of $(b, -a)$ or $0 = b - a$ or $b = a$.

5. Freedom and independence. In this section we shall establish the relations between the methods and results of the first four sections and the problems of group construction, in particular that of generation of groups by subadds. Here we say that the additive manifold M *is generated by its subadd* S, in symbols: $M = \{S\}$, if every element in M is the sum of some elements in S. If S happens to meet requirements (SR. 1) and (SR. 2) of section **4**, then $\{S\}$ is easily seen to be a group.

DEFINITION 1. *The subadd* S *of the additive manifold* M *is a free add of generators of* M [*or* M *is freely generated by* S], *if*

(a) $M = \{S\}$, *and if*

(b) *there exists to every homomorphism* $\mathfrak{s}$ *of the add* S *into some additive manifold* N *a homomorphism* $\mathfrak{t}$ *of* M *into* N *such that* $x\mathfrak{s} = x\mathfrak{t}$ *for every* x *in* S.

It is readily seen that the homomorphism $\mathfrak{t}$ is uniquely determined by $\mathfrak{s}$ and maps $\{S\}$ upon $\{S\mathfrak{s}\}$. (See Bates [1] for a discussion of free generation.)

THEOREM 1. *Suppose that* A *is an add.*

(1) *If the additive manifold* M *is freely generated by* A, *then there exists an isomorphism of* M *upon* $D(A)$ *which induces the natural homorphism in* A.

(2) *If* A *is a free add of generators of the additive manifolds* M' *and* M''.

then there exists an isomorphism of M' upon M'' which leaves invariant every element in A.

(3) *The third associative law holds in A if, and only if, A is a free add of generators of some additive manifold.*

Proof. (1) is an immediate consequence of **2**, Theorem 2 [if we let there in particular $r = 1$]; and (2) is an immediate consequence of (1). It is a consequence of **2**, Theorem 1 that $D(A)$ is freely generated by A^*; and (3) may be deduced from this fact and **3**, Corollary.

Definition 2. *The subadd S of the additive manifold M is an independent add of generators of M, if*

(a) $M = \{S\}$, *and if*

(b) $s_0 = s_1 + \cdots + s_n$ *for s_i in S and $1 < n$ imply contractability of the vector $(s_1, \cdots, s_n)$ over the add S.*

Note that the additions occurring in (b) are effected in the additive manifold M whereas the contraction of the vector $(s_1, \cdots, s_n)$ has to be effected within the add S.

Theorem 2. *If Conditions* (SR. 1) *and* (SR. 2) *of section* **4** *are satisfied by the subadd S of the additive manifold $M = \{S\}$, then the following properties of M and S are equivalent.*

(i) *S is an independent add of generators of M.*

(ii) *If $s_1, \cdots, s_n$ are elements in S such that $0 = s_1 + \cdots + s_n$, then 0 belongs to $\mathfrak{S}_S(s_1, \cdots, s_n)$.*

(iii) *M is freely generated by S and the strong associative law is satisfied by the add S.*

Proof. We note first that M is a group, as has been pointed out before. Since S is a subadd of a group, it follows from **3**, Lemma 2 that the natural homomorphism is an isomorphism; and now it is a consequence of **4**, Lemma 1 that S is strictly self-reflexive.

Assume that (i) is true and that $0 = s_1 + \cdots + s_n$. Then either $n = 1$ and $0 = s_1$ belongs to $\mathfrak{S}_S(s_1)$ or else we infer from Definition 2, (b) the existence of an integer i with the following properties: $0 < i < n$, and $s_i + s_{i+1} = t$ is an element in S. Then

$$0 = s_i + \cdots + s_{i+1} + t + s_{i+2} + \cdots + s_n;$$

and now it is clear how to prove (ii) by complete induction with respect to n (together with **1**, Lemma 1, (b)).

Assume now the validity of (ii). Since S is a subadd of the group M, the identity is an isomorphism of S into M. Hence we may infer from **2**, Theorem 1, (c) the existence of a homomorphism $\mathfrak{g}$ of $D(S)$ into M such that $x^*\mathfrak{g} = x$ for every x in S. Clearly we have $D(S)\mathfrak{g} = \{S^*\}\mathfrak{g} = \{S^*\mathfrak{g}\} = \{S\} = M$. Next we note that $D(S)$ is a group too [**4**, Proposition 1]. If w is an element in $D(S)$ such that $w\mathfrak{g} = 0$, then there exist elements $w_1, \cdots, w_m$ in S such that $w = \langle(w_1, \cdots, w_m)\rangle$; and we find that

$$0 = w\mathfrak{g} = w^*{}_1\mathfrak{g} + \cdots + w^*{}_m\mathfrak{g} = w_1 + \cdots + w_m.$$

It follows from (ii) that 0 belongs to $\mathfrak{S}_S(w_1, \cdots, w_m)$; and it follows from **3**, Lemma 1 that $(0) \sim (w_1, \cdots, w_m)$ (over S). Hence $w = 0$ and consequently $\mathfrak{g}$ is an isomorphism of $D(S)$ upon M; and now we are assured of the existence of the reciprocal isomorphism $\mathfrak{g}^{-1}$ of M upon $D(S)$. If $\mathfrak{h}$ is a homomorphism of S into the additive manifold N, then there exists a homomorphism $\mathfrak{k}$ of $D(S)$ into H such that $x^*\mathfrak{k} = x\mathfrak{h}$ for x in S (**2**, Theorem 1, (c)). Hence $x\mathfrak{h} = x^*\mathfrak{k} = x\mathfrak{g}^{-1}\mathfrak{k}$ for x in S; and the homomorphism $\mathfrak{g}^{-1}\mathfrak{k}$ of M into N induces $\mathfrak{h}$ in S. This shows that M is freely generated by S. If the vectors (0) and $(s_1, \cdots, s_n)$ are similar over S, then

$$0 = 0\mathfrak{g} = (s^*{}_1 + \cdots + s^*{}_n)\mathfrak{g} = s^*{}_1\mathfrak{g} + \cdots + s^*{}_n\mathfrak{g} = s_1 + \cdots + s_n;$$

and it follows from (ii) that 0 belongs to $\mathfrak{S}_S(s_1, \cdots, s_n)$. Hence it follows from **1**, Lemma 1 (a) that the vector $(s_1, \cdots, s_n)$ is contractable over S. Thus we have shown that the conditions (a), (b), (c) of **4**, Theorem 2 are satisfied by S; and this shows the validity of the strong associative law in S. Hence (iii) is a consequence of (ii).

Assume now the validity of (iii). Then we infer from Theorem 1, (1) the existence of an isomorphism $\mathfrak{d}$ of M upon $D(S)$ such that $x^* = x\mathfrak{d}$ for x in S. Suppose now that $1 < n$ and that the elements s_i in S satisfy $s_0 = s_1 + \cdots + s_n$ [in M]. Then we have

$$\langle(s_0)\rangle = s^*{}_0 = s_0\mathfrak{d} = s_1\mathfrak{d} + \cdots + s_n\mathfrak{d} = s^*{}_1 + \cdots + s^*{}_n$$
$$= \langle(s_1, \cdots, s_n)\rangle \text{ or } (s_0) \sim (s_1, \cdots, s_n) \text{ [over } S].$$

It follows from the strong associative law that s_0 belongs to $\mathfrak{S}_S(s_0) = \mathfrak{S}_S(s_1, \cdots, s_n)$; and the contractability of the vector $(s_1, \cdots, s_n)$ [over S] is a consequence of **1**, Lemma 1, (a). Hence S is an independent add of generators of M [Definition 2]. Thus (i) is a consequence of (iii). This completes the proof.

COROLLARY 1. *The strong associative law is satisfied by the self-reflexive add S if, and only if, S is an independent add of generators of one (and essentially only) one group.*

This is easily deduced from Theorems 1 and 2.

THEOREM 3. *If A is an independent add of generators of the additive manifold M, and if the subadd S of A is closed in A, then S is an independent add of generators of the additive submanifold $\{S\}$ of M.*

This is an almost immediate consequence of Definition 2 and 1, Definition 3.

COROLLARY 2. *If the strong associative law is satisfied by the self-reflexive add A, if the self-reflexive subadd S of A is closed in A, and if $\mathfrak{a}$ and $\mathfrak{s}$ are the natural homomorphisms of A and S respectively, then there exists an isomorphism $\mathfrak{g}$ of $D(S)$ upon $\{S\mathfrak{a}\}$ such that $x\mathfrak{s}\mathfrak{g} = x\mathfrak{a}$ for every x in S.*

This is easily deduced from the preceding results.

6. Concretions and amalgams. The following definition expresses a basic principle for the construction of adds.

DEFINITION 1. *If ϕ is a system of subadds of the add A, if every element in A belongs to at least one of the subadds in ϕ, and if the validity of the equation $a + b = c$ in A implies the existence of an add in ϕ which contains a, b, c, then A is the concretion of the adds in ϕ.*

If A is the concretion of the adds in ϕ, if B is some add, and if there is given to every add X in ϕ a homomorphism $\mathfrak{h}(X)$ of X into B such that

$$u\mathfrak{h}(X) = u\mathfrak{h}(Y) \text{ for every } u \text{ in } X \cap Y,$$

then there exists clearly one and only one homomorphism $\mathfrak{h}$ of A into B such that

$$u\mathfrak{h} = u\mathfrak{h}(X) \text{ for every } u \text{ in } X.$$

Thus homomorphisms of A are obtained by "concretion" of homomorphisms of the adds in ϕ.

The construction of adds by concretion of given adds proceeds as follows: Suppose that θ is a set of adds such that the cross cut of any two adds X and Y in θ is either vacuous or a common subadd of both X and Y. Then let $A(\theta)$ be the set of all the elements belonging to at least one of the adds in θ; and define addition in $A(\theta)$ by the rule:

$a+b=c$ if, and only if, there exists at least one X in θ which contains a, b and c and in which add X the equation $a+b=c$ is valid. (See Bates [1] for the principle of concretion.)

EXAMPLE 0. Suppose that G and H are two groups which have no elements in common. The concretion of G and H is then an add without 0, since the null-element of G cannot be added to any element in H etc. If v is a vector over the concretion C of G and H, then $\mathfrak{S}_C v$ is not vacuous if, and only if, all the coordinates of v belong to G or all the coordinates of v belong to H; and from this fact one easily deduces the validity of the strong associative law in C.

The situation indicated in Example 0 is one that we want to avoid. This leads us to the following concept.

DEFINITION 2. *If ϕ is a system of subgroups of the add A, if A is the concretion of the groups in ϕ, and if the cross cut $X \cap Y$ of any two groups X and Y in ϕ is a subgroup of both X and Y, then A is the amalgam of the groups in ϕ.*

The concretion of the groups G and H without common elements, as discussed in Example 0, is clearly not their amalgam.

EXAMPLE 1. Denote by $\{a\}$, $\{b\}$, $\{c\}$, $\{u\}$, $\{v\}$, $\{w\}$ cyclic groups, not 0, but of equal order; and let

$$U=\{a\}\oplus\{b\},\quad V=\{b\}\oplus\{c\},\quad W=\{c\}\oplus\{a\},\quad T=\{u\}\oplus\{v\}\oplus\{w\}.$$

We identify elements according to the following rule:

$$iu=i(a-b),\qquad iv=i(b-c),\qquad iw=i(c-a)\ \text{ for integral } i.$$

Then the groups U, V, W and T have the following common subgroups:

$$\begin{aligned} U\cap V&=\{b\}, & V\cap W&=\{c\}, & W\cap U&=\{a\},\\ U\cap T&=\{u\}, & V\cap T&=\{v\}, & W\cap T&=\{w\}. \end{aligned}$$

Thus we may form the amalgam A of these four groups U, V, W and T. But

the weak associative law is not satisfied by the amalgam A,

since the summation $\mathfrak{S}(a,-b,b,-c,c,-a)$ is easily seen to contain both 0 and the element $u+v+w$ in T which is not 0.

It is easily seen that amalgams are always *strictly self-reflexive adds.* It is therefore a consequence of **4**, Corollary 2 that the first and second associative laws are equivalent properties of amalgams.

EXAMPLE 2. Suppose that the group G is the direct sum of three groups H, J, K none of which is 0, in symbols: $G = H \oplus J \oplus K$. Let $U = H \oplus J$, $V = J \oplus K$, $W = K \oplus H$. Denote by A the concretion of the three groups U, V and W. Since $U \cap V$ is their common subgroup J etc., A is the amalgam of U, V and W. The identity mapping of A into G is a one-to-one homomorphism of the add A into the group G; and so it follows from **3**, Lemma 2 that the natural homomorphism is one to one. Hence A has Property (II).

Denote by h, j, k elements, not 0, in H, J and K respectively. Then $h - j$ belongs to U, $j - k$ to V and $h - k$ to W. Since $h - j$ does not belong to V nor to W, $j - k$ does not belong to W nor to U, it follows that the sum of $h - j$ and $j - k$ is undefined in the concretion A of U and V and W. But the following vector similarities are easily verified to hold over the add A:

$$(h - j, j - k) \sim (h, -j, j, -k) \sim (h, -k) \sim (h - k);$$

and this implies $(h - j)^* + (j - k)^* = (h - k)^*$, showing that the natural homomorphism of A into $D(A)$ is not an isomorphism. *Thus the amalgam A has Property* (II), *but not Property* (III).

EXAMPLE 3. Suppose that the group G is the direct sum of four groups H, J, K and L none of which is 0, in symbols: $G = H \oplus J \oplus K \oplus L$. Denote by A the concretion of the subgroups $H \oplus J$, $J \oplus K$, $K \oplus L$ and $L \oplus H$ of G. It is easily seen again that the add A is the amalgam of these four groups.

The identity mapping of A into G is clearly a one-to-one homomorphism of A into G. Suppose now that x and y are elements in A whose sum z, as defined in G, also belongs to A. Then one verifies by a simple discussion of the possible cases that x, y and z all belong to one of the four groups $H \oplus J, \cdots$; and thus it follows that the identity mapping of A into G is an isomorphism. It follows from **3**, Lemma 2 that the natural homomorphism of A into $D(A)$ is an isomorphism.

Consider now elements p, q, r, s in H, J, K and L respectively, none of which is 0. Then $p - q$, $q - r$, $r - s$ and $s - p$ belong to A, though none of the sums of any two consecutive elements (like $(p - q) + (q - r)$) is defined in A. Thus the vector $(p - q, q - r, r - s, s - p)$ is not contractable over A. On the other hand the following vector similarities hold over A:

$$(p - q, q - r, r - s, s - p) \sim (p, -q, q, -r, r, -s, s, -p) \sim (p, -p)$$
$$\sim (0).$$

Thus the strong associative law is not satisfied by the amalgam A, though the third associative law is valid in A.

We infer from **5**, Theorems 1 and 2 that this amalgam is a free add of generators of some group, though it is not an independent add of generators of any group.

What appears in the literature as "a free sum of groups with amalgamated subgroups" (Schreier [1], Neumann [1] and others) is in the present terminology a group for which a given amalgam of groups is either a free or an independent add of generators. The preceding examples illustrate some of the possibilities arising in connection with the problem of constructing "free sums of groups with amalgamated subgroups."

7. Schreier's theorem. The sole objective of this section is the proof of the following proposition.

THEOREM. *If the add A is the amalgam of the groups in the add ϕ, and if there exists a subgroup U of the add A such that $U = X \cap Y$ for any two distinct subgroups X and Y in ϕ, then the strong associative law is satisfied by the add A.*

The proof of this theorem will be effected in a number of steps.

(1) *If $\mathfrak{S}v$ is not vacuous, and if the vectors r and s are* 1-*fold contractions of v, then r and s possess either a common contraction or there exists a* 1-*fold contraction t of v such that both r, t and s, t possess common contractions.*

Assume that this is not true. Then clearly $r \neq s$. Furthermore it is impossible that $v = h + k$, $r = h' + k$, $s = h + k'$ where h' and k' are 1-fold contractions of h and k respectively, since then $h' + k'$ would be a common contraction of r and s. Thus we may assume without loss in generality that $v = h + (a, b, c) + k$, $r = h + (a + b, c) + k$, $s = h + (a, b + c) + k$. If a, b, c were in the same subgroup of A, then $h + (a + b + c) + k$ would be a common contraction of r and s. Hence a and c belong to different subgroups in ϕ; and it follows from our hypothesis that b belongs to U and that b consequently may be added to every element in A.

If there exists a 1-fold contraction h' of h, then $t = h' + (a, b, c) + k$ would be a 1-fold contraction of v; and r, t would have the common contraction $h' + (a + b, c) + k$, and s, t would have the common contraction $h' + (a, b + c) + k$. In this fashion we see that neither h nor k can have a 1-fold contraction.

If $h = h' + (d)$, then $d + a$ exists if, and only if, $d + a + b$ exists in A, since b is in U. In this case we could let $t = h' + (d + a, b, c) + k$; and t, r would have the common contraction $h' + (d + a + b, c) + k$, and t, s would have the common contraction $h' + (d + a, b + c) + k$. This is impossible; and likewise we see the impossibility of $k = (d) + k'$ together with the existence of one of the sums $c + d$ or $(b + c) + d$.

From this discussion it follows that r and s are the only 1-fold contractions of v, and that neither r nor s possesses a 1-fold contraction. But this contradicts our hypothesis that $\mathfrak{S}v$ is not vacuous [see **1**, Lemma 1, (a)]. Hence (1) is true.

(2. n) *If a is an element in $\mathfrak{S}v$, if v is a vector of length n and if w is a 1-fold contraction of v, then a belongs to $\mathfrak{S}w$.*

This proposition we prove by complete induction with respect to n [for $1 < n$]. If $v = (b, c)$, then $a = b + c$ and $w = (a)$ so that a is in $\mathfrak{S}w$ too. Hence we may assume that $2 < n$ and that (2. j) has been verified for $1 < j < n$. Suppose that v is a vector of length n, that a is in $\mathfrak{S}v$ and that w is a 1-fold contraction of v. It follows from **1**, Lemma 1, (a) that there exists a 1-fold contraction v' of v such that a belongs to $\mathfrak{S}v'$. This proves our contention if $v' = w$. If $v' \neq w$, and if v' and w possess a common contraction c, then it follows from the assertions (2. j) with $j < n$ that a belongs to $\mathfrak{S}c$; and it follows from **1**, Lemma 1, (b) that a belongs to $\mathfrak{S}w$. If v' and w do not possess a common contraction, then it follows from (1) that v possesses a 1-fold contraction t with the following properties: t and v' possess a common contraction c and t and w possess a common contraction d. As before we verify successively that a belongs to $\mathfrak{S}c$, hence to $\mathfrak{S}t$, hence to $\mathfrak{S}d$ and therefore to $\mathfrak{S}w$. This completes the proof of (2. n).

From (2. n) and **1**, Lemma 1, (b) we deduce that $\mathfrak{S}v = \mathfrak{S}w$ whenever w is a contraction of v; or in the terminology of section **2**, (S. 1) whenever v and w are directly similar. But now one verifies immediately [see **2**, (S. 2)] that $v \sim w$ implies $\mathfrak{S}v = \mathfrak{S}w$. Hence the strong associative law holds in A, as we claimed.

Remark 1. It follows from **5**, Corollary 1 that the assertion of the preceding theorem is equivalent to the fact that A is an independent add of generators of some group G; and it is easily seen that G is, in the terminology of Schreier [1], "a free sum of groups with *one* amalgamated subgroup." Thus we have obtained a new proof of Schreier's theorem. It would, of course, have been possible to obtain a proof of the preceding theorem by direct reference to Schreier's original theorem and **5**, Corollary 1; but it seemed appropriate to design a proof based on direct verification of the strong associative law.

COROLLARY. *The strong associative law is satisfied by every amalgam of two groups.*

This is obviously a special case of the preceding theorem.

Remark 2. In the subsequent applications we shall only use this Corollary; and the preceding theorem will be recovered as a special case of some of the later results which are derived from this corollary. But the proof of our theorem would not have been simplified if we had only considered amalgams of two groups.

8. The structure of the amalgamations. If the add A is the amalgam of the groups in the set ϕ, and if X and Y are two different groups in ϕ, then the intersection $X \cap Y$ may be considered as having arisen by amalgamation of a certain subgroup of X with a certain subgroup of Y. The examples of Section **6** and Schreier's theorem show that the question whether or not the strong associative law is satisfied by some amalgam will depend on the structure of these amalgamations; and the present section is devoted to a study of this question.

LEMMA 1. *Suppose that the add A is the concretion of the group G and the self-reflexive add R, that R and G and A have the same null-element, that $R \cap G$ is an independent add of generators of the group G, and that the strong associative law is satisfied by A. Then the strong associative law is satisfied by R.*

Proof. It follows from our hypotheses that **4**, (SR. 1) and **4**, (SR. 2) are satisfied by the add A. Hence we deduce from the strong associative law and **4**, Lemma 1 that A is strictly self-reflexive. Consequently $-x$ is uniquely determined by x; and now it is easy to see that R and $R \cap G$ are strictly self-reflexive too.

It is a consequence of the strong associative law that A is an independent add of generators of an essentially uniquely determined group H (**5**, Corollary 1). Since R is a subadd of A, and A is a subadd of the group H, R itself is a subadd of the group H. Denote by K the subgroup of H, generated by R. Since R is self-reflexive, every element in K is a sum of elements in R or $K = \{R\}$. We prove:

(1. a) *R is an independent add of generators of the group K.*

Suppose that $2 \leq n$, that $r_1, \cdots, r_n$ are elements in R and that $0 = r_1 + \cdots + r_n$. We want to prove the contractability of the vector $(r_1, \cdots, r_n)$ over R. If this were not true, then $r_i + r_{i+1}$ for $0 < i < n$ would not belong to R. We have to show that this hypothesis leads to a contradiction. We note first the existence of some r_i that does not belong to G. For otherwise $(r_1, \cdots, r_n)$ would be a vector over $R \cap G$. But $R \cap G$ is an independent add of generators of G; and the equation $0 = r_1 + \cdots + r_n$ would hold in G too and would imply the contractability of the vector $(r_1, \cdots, r_n)$ over $R \cap G$ and hence over R. Consequently there exist integers $n(j)$ with the following properties:

$$1 \leq n(1) < n(2) < \cdots < n(k) \leq n;$$

r_i belongs to $R \cap G$ if, and only if, $i \neq n(j)$ for every j.

Now we define:

If $1 < n(1)$, then $g(0) = r_1 + \cdots + r_{n(1)-1}$;

if $n(i) + 1 < n(i+1)$, then $g(i) = r_{n(i)+1} + \cdots + r_{n(i+1)-1}$;

if $n(k) < n$, then $g(k) = r_{n(k)+1} + \cdots + r_n$.

It follows from our choice of the integers $n(j)$ that $g(i)$ is either undefined or an element in G. Since the sum of the r_i (in their natural order) is 0, and since addition in H is associative, the sum of the $r_{n(j)}$ and $g(j)$ in their natural order is 0 too. Indicating typical terms only this last sum has the following form:

$$(*) \quad 0 = \cdots + r_{n(i)} + r_{n(i+1)} + \cdots + r_{n(j)} + g(j) + r_{n(j+1)} + \cdots$$

(where $n(i) + 1 = n(i+1)$ and $n(j) + 1 < n(j+1)$.)

We note that the $r_{n(j)}$ belong to R, but not to G, whereas the $g(j)$ belong to G. Hence all the summands of (*) belong to the add A. But A is an independent add of generators of the group H; and so there exist two consecutive summands in the sum (*) whose sum belongs to A. Consequently there arise three possibilities:

Case 1. There exists an i such that $n(i) + 1 = n(i+1)$ and such that $r_{n(i)} + r_{n(i+1)}$ belongs to A.

Since by hypothesis the sum of these two elements does not belong to R, and since A is the concretion of R and G, this would imply that $r_{n(i)}, r_{n(i+1)}$ and $r_{n(i)} + r_{n(i+1)}$ belong to G. But this contradicts our choice of the integers $n(j)$. Thus this case cannot occur.

Case 2. There exists an integer i such that $g(i-1)$ and $r_{n(i)}$ are both defined and such that $g(i-1)+r_{n(i)}$ exists in A.

Since $r_{n(i)}$ does not belong to G, and since A is the concretion of R and G, it follows that $g(i-1)$, $r_{n(i)}$ and $g(i-1)+r_{n(i)}$ all belong to R. Thus $g(i-1)$ belongs to $R\cap G$. But

$$g(i-1)=r_{n(i-1)+1}+\cdots+r_{n(i)-1}$$

(letting $n(0)=0$, if necessary) where $g(i-1)$ as well as the summands of the right side belong to $R\cap G$. It is impossible that $n(i-1)+1=n(i)-1$, since then $g(i-1)=r_{n(i)-1}$ and consequently $r_{n(i)-1}+r_{n(i)}$ would be an element in $R\cap G$. Thus the sum defining $g(i-1)$ contains more than one summand. However, $G\cap R$ is supposed to be an independent add of generators of G. Consequently the defining equation for $g(i-1)$ implies the contractability of the vector $(r_{n(i)1+1},\cdots,r_{n(i)-1})$ over $R\cap G$. Thus the sum of two consecutive r's belongs to $R\cap G$; and we have obtained a contradiction again. Hence this case cannot occur.

Case 3. There exists an integer i such that $r_{n(i)}$ and $g(i)$ are both defined and such that $r_{n(i)}+g(i)$ belongs to A.

The impossibility of this case 3 is shown exactly as that of case 2.

Thus we have shown: If $r_1,\cdots,r_n$ are elements in R such that $1<n$ and $0=r_1+\cdots+r_n$, then the vector $(r_1,\cdots,r_n)$ is contractable over R. From this fact one deduces easily the validity of Condition (ii) of **5**, Theorem 2. Hence (1. a) is true and the strong associative law is satisfied by R, as we claimed.

LEMMA 2. *Suppose that the add A is the concretion of the group G and the self-reflexive add R, that $R\cap G$ is a subgroup of G. Then the strong associative law holds in A if, and only if, it is satisfied by R.*

Proof. We note first the fairly obvious facts that A is self-reflexive and that R is closed in A. (For if a, b are elements in R, $a+b=c$ in A, then either a, b, c belong to R, as we want to show, or a, b, c belong to G, since A is the concretion of R and G. But in the latter case a, b and hence $a+b=c$ belong to the subgroup $R\cap G$ of G, so that $a+b=c$ belongs to R in either case.)

If the strong associative law holds in A, then it follows from **3**, Proposition 1, (b) that the strong associative law holds in the closed subadd R of A.

Assume conversely the validity of the strong associative law in R. Then we infer from **5**, Corollary 1 the existence of one and essentially only one group H such that R is an independent add of generators of H. Since the intersection of R and G is the subgroup $S = R \cap G$ of G, we may assume without loss in generality that

(1) $S = R \cap G = H \cap G$.

We form the concretion K of the two groups H and G, subject to the rule (1). We show

(2) A is a subadd of K such that $A \cap H = R$.

It is clear that A is contained in K. Consider elements a, b, c in A. If $a + b = c$ holds in A, then either a, b, c are in R and $a + b = c$ holds in R and hence in H and in K or else a, b, c are in G and $a + b = c$ holds in G and hence in K. If conversely $a + b = c$ holds in K, then either a, b, c are in G and $a + b = c$ holds in G and hence in A or else a, b, c are in H and $a + b = c$ holds in H. But if the elements a, b, c in H belong to A, then they belong to the subadd R of H and so $a + b = c$ holds in R and hence in A. This proves (2).

(3) K is the concretion of A and H.

It is clear that every element in K belongs to A or to H, because every element in K belongs already to $G \leq A$ or to H. Suppose now that a, b, c are elements in K such that $a + b = c$. Then either a, b, c are in H and $a + b = c$ holds in H or else a, b, c are in G and $a + b = c$ holds in $G \leq A$, since K is the concretion of G and H.

(4) The strong associative law is satisfied by K.

This is a consequence of Schreier's theorem (**7**, Corollary), since, by (1), K is the amalgam of two groups H and G.

Since the subadd $A \cap H = R$ of K is an independent add of generators of the group H, it follows from (1) to (4) that we may apply Lemma 1, showing that the strong associative law is satisfied by A. This completes the proof.

PROPOSITION 1. *Suppose that the add A is the amalgam of the finitely many groups $G_1, \cdots, G_n$ meeting the following requirements:*

(a) *The join $A(i)$ of $G_1, \cdots, G_i$ is their concretion.*

(b) *$A(i) \cap G_{i+1}$ is a subgroup of G_{i+1}.*

Then the strong associative law is satisfied by A.

Proof. Note that every $A(i)$ is the amalgam of $G_1, \cdots, G_i$, since A is the amalgam of all the G_i, and since $A(i)$ is the concretion of $G_1, \cdots, G_i$. Note furthermore that $A(1) = G_1$ and $A(n) = A$.

We are going to prove by complete induction with respect to i that every $A(i)$ satisfies the strong associative law. This is certainly true for the group $A(1)$. Thus we assume that $1 < i$ and that the strong associative law holds in $A(i-1)$. It follows from (a) that $A(i)$ is the concretion of the group G_i and of the self-reflexive add $A(i-1)$; and it follows from (b) that the cross cut of $A(i-1)$ and G_i is a subgroup of G_i. It follows from Lemma 2 that the validity of the strong associative law in $A(i-1)$ implies its validity in $A(i)$. This completes the induction. Consequently the strong associative law is satisfied in particular by $A(n) = A$.

Remark 1. The need for condition (a) of Proposition 1 arises from the following fact. Suppose that a, b, c are elements in G_1, G_2, G_3 respectively; and suppose that $a + b = c$. The fact that A is the amalgam of $G_1, \cdots, G_n$ implies the existence of some G_i which contains a, b, c such that $a + b = c$ holds in G. But we are not assured that i may be chosen from $1, 2, 3$ unless we impose condition (a).

It is obvious how to deduce criteria for the validity of the strong associative law by a formal combination of Proposition 1 and **3**, Proposition 2. More interesting than a general criterion of this type seem some special instances.

THEOREM 1. *Suppose that the add A is the amalgam of the groups in ϕ and that there exists a group U in ϕ with the following property:*

(*) *If X, Y, Z are three distinct groups in ϕ such that $X \cap Y$ is neither part of U nor of Z, then $Z \cap Y \leq Y \cap X$.*

Then the strong associative law is satisfied by A.

Proof. Suppose that $G_1, \cdots, G_n$ are a finite number of groups in ϕ which are all different from U. Denote by J the set theoretical join of U, $G_1, \cdots, G_n$. We deduce a number of properties of J.

(1.1) *If X is a group in ϕ, then $X \cap J = X \cap U$ or else there exists an integer m such that $1 \leq m \leq n$ and $X \cap J = X \cap G_m$.*

This proposition is certainly true, if X is one of the G_i. Hence we may assume that $X \neq G_i$ for every i. The proposition is likewise true, if $X \cap G_i \leq U$ for every i. Thus we may assume now without loss in generality that there exists an integer k with the following properties:

$$0 \leq k < n;\ X \cap G_i \leq U \text{ for } i \leq k;\ X \cap G_i \nleq U \text{ for } k < i \leq n.$$

We prove:

(a) If $k < i$, then $U \cap X \leq X \cap G_i$.

This follows by applying (*) on the triplet G_i, X, U in ϕ, since $G_i \cap X$ is not part of U.

(b) If $k < i$ and $k < j$, then $G_i \cap X \leq X \cap G_j$ or $G_j \cap X \leq X \cap G_i$.

If $G_i \cap X \nleq X \cap G_j$, then $G_i \cap X$. is neither part of U nor of G_j. Thus we may apply (*) on the triplet G_i, X, G_j; and it follows that $G_j \cap X \leq X \cap G_i$. This proves (b).

It follows from (b) that the subgroups $X \cap G_i$ of X with $k < i \leq n$ form a finite ordered set. Hence there exists an integer m such that $k < m \leq n$ and such that $k < i \leq n$ implies $X \cap G_i \leq X \cap G_m$. It follows from (a) that $U \cap X \leq X \cap G_m$. If $i \leq k$, then it follows from our choice of k that $X \cap G \leq X \cap U$; and this implies $X \cap G_i \leq X \cap G_m$. Combining all these facts it follows that $X \cap J = X \cap G_m$; and this completes the proof of (1.1).

(1.2) *J is the concretion of* $U, G_1, \cdots, G_n$.

Suppose that x, y, z are elements in J such that $x + y = z$ holds in A. Then there exists a group X in ϕ which contains x, y, z such that $x + y = z$ holds in X. It follows from (1.1) that either $X \cap J = X \cap U$ or else $X \cap J = X \cap G_m$ for some m. In the first case x, y, z are in U and $x + y = z$ holds in U; and in the second case x, y, z are in G_m and $x + y = z$ holds in G_m. This proves (1.2). (Note that it is possible to prove by quite similar arguments the fact that J is closed in A.)

Suppose now that F is a finite subset of A. Then there exists a finite number of groups $P_1, \cdots, P_p$ in ϕ which are all different from U such that F is part of the join of $U, P_1, \cdots, P_p$. Denote by $J(i)$ the join of U, $P_1, \cdots, P_i$. Then $J(0) = U$ and $F \leq J(p)$. It follows from (1.2) that $J(i)$ is the concretion of $U, P_1, \cdots, P_i$; and it follows from (1.1) that $J(i) \cap P_{i+1}$ is a subgroup of P_{i+1}—remember that A is the amalgam of the groups in ϕ. It follows from Proposition 1 that the strong associative law is satisfied by $J(p)$. Consequently it follows from **3**, Proposition 2 that the strong associative law is satisfied by A.

Remark 2. Since the antecedent in condition (*) is symmetric in X and Y, this condition implies actually the following stricter statement:

If X, Y, Z are three distinct groups in ϕ such that $X \cap Y$ is neither part of U nor of Z, then $Z \cap Y \leq Y \cap X$ and $Z \cap X \leq X \cap Y$. But from these inequalities one deduces readily that

$$Z \cap X = X \cap Y \cap Z = Z \cap Y.$$

The following extreme special case of Theorem 1 is not only of great interest in itself, but also convenient for applications. (Another extreme special case will be treated in section **9** below.)

COROLLARY 1. *Suppose that the add A is the amalgam of the groups in ϕ and that there exists a subgroup U of the add A with the following property*:

(**) *If X and Y are different groups in ϕ, then $X \cap Y \leq U$.*

Then the strong associative law is satisfied by A.

Proof. It is easily verified that A is the amalgam of the groups in ϕ together with U; and thus we may assume without loss in generality that U belongs to ϕ. It is clear that condition (*) of Theorem 1 is satisfied by U and ϕ; and thus the strong associative law holds in A.

Remark 3. If we change in condition (**) inequality into equality, we obtain just Schreier's theorem (section **7**) which thus reappears as a special case of Theorem 1. We note that only a special instance of Schreier's theorem (amalgam of two groups) has been used in our deductions.

Our next application is mainly devoted to obtaining a theorem which, on the basis of **5**, Theorem 1, is equivalent to a theorem discovered by H. Neumann [1]. For its formulation some concepts are needed.

Suppose that the add A is the amalgam of the groups in the set ϕ. If X is a given group in ϕ, then we denote by (X, ϕ) *the subgroup of X which is generated by all the subgroups $X \cap Y$ of X for Y in ϕ, $Y \neq X$*; and we denote by (A, ϕ) *the set of all the elements in A which belong to at least one (X, ϕ)*.

LEMMA 3. *If A is the amalgam of the groups in ϕ, then $A(\phi)$ is the amalgam of the groups (X, ϕ); and (A, ϕ) is closed in A.*

Proof. Suppose that x, y, z are elements in A such that $x + y = z$; and suppose that x and y belong to (A, ϕ). Since A is the amalgam of the groups in ϕ, there exists a group G in ϕ such that $x + y = z$ holds in G. Since x is in (A, ϕ), there exists an X in ϕ such that x is in the subgroup (X, ϕ).

If $X \neq G$, then x belongs to the subgroup $(X, \phi) \cap G$ of $X \cap G$ which in turn is a subgroup of (G, ϕ). Thus x belongs necessarily to the subgroup (G, ϕ) of G; and the same holds for y. Hence $x + y = z$ belongs to (G, ϕ) too. This shows that (A, ϕ) is closed in A and that (A, ϕ) is the concretion of the subgroups (X, ϕ). If $X \neq Y$, then $(X, \phi) \cap (Y, \phi) = X \cap Y$ which is a common subgroup of $X, Y, (X, \phi)$ and (Y, ϕ); and thus (A, ϕ) is the amalgam of the groups (X, ϕ) for X in ϕ.

THEOREM 2. *Suppose that the add A is the amalgam of the groups in ϕ.*

(a) *The natural homomorphism of A into its derived manifold is an isomorphism if, and only if, the natural homomorphism of (A, ϕ) into its derived manifold is an isomorphism.*

(b) *The strong associative law holds in A if, and only if, it is satisfied by (A, ϕ).*

Proof. It is a consequence of Lemma 3 that (A, ϕ) is closed in A. Hence it follows from **3**, Proposition 1 that (A, ϕ) has Property (III) or (IV) whenever A has the corresponding property, showing the necessity of our conditions.

If the strong associative law holds in (A, ϕ), then the natural homomorphism of (A, ϕ) into its derived manifold is an isomorphism. Thus it follows from **5**, Theorem 1, (c) that (A, ϕ) is a free add of generators of one and essentially only one group G (see also **4**, Proposition 1) whenever either the third or the fourth associative law is satisfied. We note that it will suffice for the first part of our proof to assume that

(1) *(A, ϕ) is a subadd of the group G;*

and only during the second part of our proof will there be restrictions imposed on our choice of the group G.

We may assume without loss in generality that

(2) $G \cap A = (A, \phi)$.

With this condition (2) in mind we form the concretion K of A and G. Let θ be the set of groups composed of G and the groups in ϕ. Since $G \cap X = (X, \phi)$ for X in ϕ, it follows that

(3) *K is the amalgam of the groups in θ.*

If the elements a, b, c belong to A, and if $a + b = c$ holds in K, then either a, b, c belong to A and $a + b = c$ holds in A; or else a, b, c belong to

G and $a + b = c$ holds in G. But in the latter case a, b, c belong to $G \cap A = (A, \phi)$. Since (A, ϕ) is a subadd of G, it follows that $a + b = c$ holds in (A, ϕ) and hence in A; and now it is easy to see that

(4) *A is a subadd of K.*

If X and Y are two different groups in θ, then it follows from our construction of G that $X \cap Y \leq G$; if neither X nor Y is G, then $X \cap Y \leq (A, \phi) \leq G$; thus (**) of Corollary 1 is satisfied by the amalgam K of the groups in θ and we infer from Theorem 1 that

(5) *the strong associative law holds in K.*

From (5) we infer the validity of Property (III) in K and in all its subadds (**3**, Proposition 1. (a)). Thus the subadd A of K has Property (III), and this proves the validity of (a).

Assume now the validity of the strong associative law in (A, ϕ). Then it follows from **5**, Corollary 1 that (A, ϕ) is an independent add of generators of one and essentially only one group; and thus we may assume now that (A, ϕ) is an independent add of generators of the group G. Now it follows from (2), (4), (5) and Lemma 1 that the strong associative law holds in A too; and this completes the proof.

Remark 4. It is not difficult to see that Theorem 1 is a special case of Theorem 2, (b).

Remark 5. Using **5**, Theorem 1 and **5**, Corollary 1 it is possible to restate Theorem 2 in the following form:

(a′) The amalgam A of the groups in ϕ is a free add of generators of some group if, and only if, (A, ϕ) is a free add of generators of some group.

(b′) The amalgam A of the groups in ϕ is an independent add of generators of some group if, and only if, (A, ϕ) is an independent add of generators of some group.

Proposition (a′) is equivalent to Neumann's theorem (see H. Neumann [1], p. 599, Theorem).

9. Homogeneous Amalgams.

Homogeneity of amalgams may be defined in a variety of ways. We consider only a fairly strong type of homogeneity.

DEFINITION. *If the add A is the amalgam of the groups in the set ϕ, and if, for every non vacuous subset θ of ϕ, the join of the groups in θ is a closed subadd of A, then A is a homogeneous amalgam of the groups in ϕ.*

It is trivial to construct amalgams which are not homogeneous. The following characterization of homogeneous amalgams will be convenient to use.

LEMMA. *The following properties of the amalgam A of the groups in the set ϕ are equivalent.*

(i) *A is the homogeneous amalgam of the groups in ϕ.*

(ii) *If x, y, z are elements in A such that $x + y = z$, and if x belongs to X in ϕ and y to Y in ϕ, then x, y, z belong all to X or they all belong to Y.*

(iii) *If X, Y, Z belong to ϕ, then $X \cap Y \leq Y \cap Z$ or $Z \cap Y \leq Y \cap X$.*

Proof. Assume first the validity of (i). Suppose that x, y, z are elements in A, that $x + y = z$, that x is in X, y is in Y. The join J of the groups X and Y is closed in A; and so z belongs to J too. This implies that z belongs to X or to Y; and we may assume without loss in generality that z belongs to X. There exists a group Z in ϕ such that $x + y = z$ holds in Z, since A is the amalgam of the groups in ϕ. Consequently x and z belong to the common subgroup $X \cap Z$ of X and of Z. This implies that y belongs to $X \cap Z$ too. Thus x, y, z belong to X. Hence (ii) is a consequence of (i).

Assume next the validity of (ii). Consider groups X, Y, Z in ϕ; and assume that $X \cap Y \not\leq Y \cap Z$ and $Z \cap Y \not\leq Y \cap X$. Then there exists an element x in $X \cap Y$ which does not belong to $Y \cap Z$; and there exists an element z in $Z \cap Y$ which does not belong to $Y \cap X$. Thus x and z are both in Y; and consequently there exists a well-determined element y in the group Y such that $z + x = y$. On the other hand x cannot be in Z nor z in X, contradicting (ii). This shows that (iii) is a consequence of (ii).

Assume finally the validity of (iii). Suppose that x, y, z are elements in A such that $x + y = z$; and suppose that they belong to the groups X, Y and Z respectively (in ϕ). There exists a group W in ϕ such that $x + y = z$ holds in W, since A is the amalgam of the groups in ϕ. It follows from (iii) that $X \cap W \leq W \cap Y$ or $Y \cap W \leq W \cap X$. Since x belongs to $X \cap W$, it follows in the first of these cases that x and y belong to the common subgroup $W \cap Y$ of Y and W, implying that x, y, z belong to $W \cap Y \leq Y$; and in the same fashion we show in the second case that x, y, z belong to X. But from this fact it is easily deduced that A is the homogeneous amalgam of the groups in ϕ. Thus (i) is a consequence of (iii).

THEOREM. *Suppose that A is the homogeneous amalgam of the groups in the set ϕ.*

(a) *If θ is a nonvacuous subset of ϕ, then the join $A(\theta)$ of the groups in θ is the homogeneous amalgam of the groups in θ.*

(b) *The strong associative law is satisfied by every $A(\theta)$ (and in particular by A).*

(c) *If A is an independent add of generators of the group G, then $A(\theta)$, for θ a nonvacuous subset of ϕ, is an independent add of generators of the subgroup $\{A(\theta)\}$ of G.*

Proof. (a) is an immediate consequence of the preceding Lemma; (b) may be deduced from the preceding Lemma in conjunction with **8**, Theorem 1; and (c) is a consequence of **5**, Theorem 3, since $A(\theta)$ is closed in A.

UNIVERSITY OF ILLINOIS,
URBANA, ILLINOIS.

BIBLIOGRAPHY.

E. Artin [1] "Das freie Produkt von Gruppen," F. Klein: *Vorlesungen über höhere Geometrie.* Herausgegeben von W. Blaschke; Berlin, 1926, § 92, pp. 361-363.

[2] "The free product of groups," *American Journal of Mathematics*, vol. 69 (1947), pp. 1-4.

Grace Bates [1] "Free nets and loops and their generalizations," *American Journal of Mathematics*, vol. 69 (1947), pp. 499-550.

H. Neumann [1] "Generalized free products with amalgamated subgroups," *American Journal of Mathematics*, vol. 70 (1948), pp. 590-625.

O. Schreier [1] *Die Untergruppen der freien Gruppen.* Hamburger Abhandlungen, vol. 5 (1927), pp. 161-183.

THE ISOPERIMETRIC PROBLEM FOR MINKOWSKI AREA.*

By HERBERT BUSEMANN.

The present paper centers about the following problem: A positive continuous function $\sigma(u)$ is defined on all unit vectors u of E^n. *To find among all simple closed surfaces D in a certain class Γ and with a given "area" $\int_D \sigma(u)\,dS$ those which bound the maximal (Euclidean) volume $V(D)$.* Here dS means the (Euclidean) surface element of D and u the exterior normal of D "at dS."

For the rather restricted class Γ which consists of all convex surfaces and the non-convex surfaces whose first partials piecewise satisfy Lipschitz conditions, the problem is solved by the *isoperimetric inequality*

$$(*) \qquad \int_D \sigma(u)\,dS \geq nV^{(n-1)/n}(D)\,V^{1/n}(K)$$

where K is the polar reciprocal (with respect to the unit sphere) to the boundary of the convex closure of the surface C: $\sigma^{-1}(u)u$; and the equality sign holds only for D homothetic to K.[1]

The author's interest in the problem is due to the fact that (*) *solves the isoperimetric problem for the intrinsic area in Minkowski (or finite dimensional Banach) spaces* (Section **6**). *The solution is in general not a Minkowski sphere, but a new convex surface which proves of great importance for the theory of Finsler spaces.*[2] The reader interested in these questions is referred to the author's lecture on "The Geometry of Finsler Spaces" which is going to appear in the Bulletin of the American Mathematical Society. The present paper shows only that Minkowski's definition of area in Euclidean spaces by means of parallel sets carries over to Minkowski spaces if the parallel sets are formed with solutions of the isoperimetric problem instead of spheres.

---

* Received February 11, 1948; revised January 31, 1949.

[1] If u means the interior normal, then K must be replaced by the surface symmetric to K with respect to the origin.

[2] For the plane the problem was solved in [3] and the methods of [3] are used here in Sections 1 and 5. However, in the plane case the solution of the isometric problem reduces after a rotation through $\pi/2$ to the sphere of the dual space (in Banach space terminology), and is therefore not an essentially new geometric object.

Using (*) presupposes the validity of the representation $\int \sigma(u)\,dS$ for the intrinsic area of a surface of class Γ in Minkowski spaces. But [2], where this area is defined, establishes this representation only for surfaces of class D', and therefore not for all convex surfaces. The final Section **7** of the present paper fills this gap by proving the much more general theorem that *$\int \sigma(u)\,dS$ represents the intrinsic area of any rectifiable manifold in a Finsler space.* For rectifiable manifolds in Riemann spaces $\int \sigma(u)\,dS$ also represents the various other areas (of Lebesgue, Radó, Federer, . . .) none of which is intrinsic by definition. Thus the present theorem contributes to the theory of these areas the information that they are intrinsic in the rectifiable case.

The form of (*) indicates a close connection with the Brunn-Minkowski Theory, in fact, the area $\int \sigma(u)\,dS$ was first studied by Minkowski in [9, § 27], and (*) follows for convex C and D directly from Minkowski's inequality (Section **1**). For non-convex D the relation (*) follows from Lusternik's generalization [8] of the Brunn-Minkowski Theorem to non-convex sets, see Section **2**. However, Lusternik's method does not yield the condition for the equality sign in (*). Therefore Dinghas and Schmidt proved in [5] Lusternik's result again for the special case where C is a sphere, by adapting Kneser and Süss' proof of the Brunn-Minkowski Theorem (compare [1, pp. 88-90]), and showed how this method leads through an elegant additional device to the conditions for the equality sign in (*).

Sections **3** and **4** of the present paper use the method of [5], at times verbally, for arbitrary convex C instead of spheres, and go slightly farther, to yield *a necessary and sufficient condition for the equality sign in the isoperimetric problem for the Minkowski area corresponding to C.*

This result implies (*) for convex C. The passage to arbitrary C (Section **5**) is accomplished in a surprisingly simple manner by an idea which the author called regularity principle,[3] and which seems applicable to many similar problems.

1. Let $\sigma(u)$ be a positive continuous function defined on the unit vectors u of E_n, and such that the surface $C\colon \sigma^{-1}(u)u$ is convex. The definition of $\sigma(u)$ is extended to arbitrary vectors x by

$$\sigma(x) = |x|\,\sigma(x/|x|) \text{ for } x \neq 0, \qquad \sigma(x) = 0 \text{ for } x = 0.$$

Then C has the equation $\sigma(x) = 1$ and $\sigma(x)$ is a convex positive homogeneous function of the first degree in x. Therefore it is supporting function of a convex body L with boundary K. It is well known [9, § 8] that K originates from C by a polar reciprocity in the unit sphere $|x| = 1$. If $\sigma(x)$ is analytic and D is any other analytic closed convex surface bounding the convex body E, then

---

[3] Compare [4]. The author was incautious enough to state in [4], that the present problem for non-convex D was hopeless. He was then not aware of the paper [8] by Lusternik.

$$\int_D \sigma(u)\,dS = nV(E, L, \cdots, L), \tag{1}$$

where (see [9, § 27] or [1, p. 64]) u is the exterior normal of D "at dS" and

$$V_1(E, L) = V(E, L, \cdots, L) \tag{2}$$

is the mixed volume of E and $(n-1)$-times L. If D and K are arbitrary closed convex surfaces, then sequences $\{D_\nu\}$ and $\{K_\nu\}$ of analytic convex surfaces exist, which tend to D and K respectively ([1, section 27]). Convergence of convex surfaces implies convergence of their supporting planes. Since $V_1(D, K)$ depends also continuously on D and K ([1, p. 40]) it follows from (1) and (2) that for any convex D and K

$$\int \sigma(u)\,dS = nV_1(E, L) = \lim \rho^{-1}(|E + \rho L| - |E|). \tag{3}$$

The integral on the left is extended over all points of D where the tangent plane exists, $|X|$ denotes generally the exterior Lebesgue measure of the set X, and $X + \rho Y$ means the set consisting of all points of the form $x + \rho y$ with $x \varepsilon X$ and $y \varepsilon Y$. For the last part of (3) compare [1, p. 47]. The number $nV_1(E, L)$ is called the area of D relative to K and was first studied by Minkowski ([9, § 27]). The inequality of Minkowski (see [1, p. 91] or the next section of this paper)

$$V_1{}^n(E, L) \geqq |E|^{n-1}|L|$$

and (3) yield

$$\int_D \sigma(u)\,dS \geqq n\,|E|^{(n-1)/n}\,|L|^{1/n} \tag{4}$$

and the equality sign holds only for E homothetic to L or D homothetic to K. Therefore

(5) *If $C: \sigma^{-1}(u)u$ is convex then the surfaces homothetic to the polar reciprocal K of C and only these minimize $\int_D \sigma(u)\,dS$ among all convex surfaces D with a given volume $|E|$.*

2. The discussion of the same problem for non-convex D rests on the

expression which the right side of (3) gives for $\int_D \sigma(u)\,dS$. If H is the interior of K, then for any convex E

$$| E + \rho H | = | E + \rho L | .$$

Denoting the closure of the set X by $\bar{X}$, any set X satisfies the relation

$$X + \rho H = \bar{X} + \rho H. \tag{6}$$

Therefore we define generally as Minkowski area $B(M)$ relative to K of the boundary of a set X with $| \bar{X} | < \infty$, the number

$$B(X) = \liminf_{\rho \to 0+} \rho^{-1}(| X + \rho H | - | \bar{X} |). \tag{7}$$

Because of (6)

$$B(X) = B(\bar{X}). \tag{8}$$

For this area the isoperimetric problem can be solved *completely.* In order to formulate the solution adequately we introduce the following notation. For any bounded set M with $| \bar{M} | > 0$ and a given direction u, let a' be the least upper bound of those d for which the intersection of $\bar{M}$ with the set $xu \leq d$ has measure 0 and b' the greatest lower bound of the d for which $\bar{M}$ intersects $xu \geq d$ is a set of measure 0. We call $a' \leq xu \leq b'$ the *essential supporting strip* $W(u)$ of M normal to u. Obviously $W(u) = W(-u)$ and $| M - M \cap W(u) | = 0$. The intersection

$$M_E = M \cap \prod_u W(u)$$

is called the *essential part of* M. By definition $\overline{M_E} = (\bar{M})_E$. There is a countable set $\{u_\nu\}$ of directions u such that $\prod_u W(u) = \prod_\nu W(u_\nu)$. Since

$$M - M_E = M - M \cap \prod W(u_\nu) = \bigcup_\nu [M - M \cap W(u_\nu)]$$

it follows that

$$| M - M_E | = 0. \tag{9}$$

Corresponding to (4) we are going to prove the isoperimetric inequality:

THEOREM 1. *For any set* M *with* $0 < | \bar{M} | < \infty$,

$$B(M) \geq n | \bar{M} |^{(n-1)/n} | L |^{1/n}, \tag{10}$$

and the equality sign holds if and only if $\bar{M}_E$ *is homothetic to* L *and* $B(M) = B(M_E)$.

Notice that the theorem is trivial for unbounded M since then $|M+\rho H|=\infty$ for any ρ, so that $B(M)=\infty$. In the proof we assume therefore that M is bounded. We call topological sphere a topological image D of the unit sphere $|x|=1$. If M is the interior of D then $M_E=M$. Therefore Theorem 1 yields .

(11) *If C is convex then the surfaces homothetic to K and only these minimize $B(M)$ among all topological spheres D with a given (positive, finite) volume $|M|$ or $|\bar{M}|$, where M is the interior of D.*

(10) without the discussion of the equality sign follows from the Brunn-Minkowski Theorem $|M+\rho H|^{1/n}\geq|\bar{M}|^{1/n}+\rho|L|^{1/n}$ which yields

$$|M+\rho H|-|\bar{M}|\geq n\rho|\bar{M}|^{(n-1)/n}\cdot|L|^{1/n}.$$

The relation $|X+\rho R|^{1/n}\geq|X|^{1/n}+\rho|Y|^{1/n}$ was given for arbitrary measurable X and Y by Lusternik in [8]. For reasons explained in the introduction we follow very closely the method of [5] to establish Theorem 1 and the Brunn-Minkowski Theorem for arbitrary convex Y. With minor modifications the present proof works for any Y which contains an open set Z with $|Z|=|Y|$.

3. The purpose of the next two sections is then to prove simultaneously Theorem 1 and

THEOREM 2. *If M is a bounded set and K' is a bounded convex set with interior points, then for any $\rho>0$*

$$|M+\rho K'|^{1/n}\geq|\bar{M}|^{1/n}+\rho|K'|^{1/n} \tag{12}$$

and the equality sign holds if and only if $\bar{M}$ is homothetic to $\bar{K}'$.

Note. If either M or K' is not bounded or K' has no interior points (12) is trivially correct. The conditions for the equality sign are different, but equally trivial.

If H is the set of interior points of K' then $|H|=|K'|$, hence Theorem 2 follows, because of (6), from the special case where $K'=H$ and $M=\bar{M}$. Instead of (12) we therefore prove

$$|M+\rho H|^{1/n}\geq|M|^{1/n}+\rho|H|^{1/n} \text{ for closed } M. \tag{13}$$

Theorem 2 is trivial for $|M|=0$. For if M contains more than one point then $|M+\rho H|>\rho|H|$. And if M consists of one point, then $|M+\rho H|=\rho|H|$ and M is homothetic to H. Therefore we may assume that $|M| > 0$, and it means no restriction if we assume in addition that $|M| = |H|$, because the general case follows immediately from this case.

The proof proceeds by *induction.* The very simple case $n = 1$ is left to the reader, and Theorem 2 is assumed to hold for $n - 1$. In E_n introduce rectangular coordinates $x, y_1, \cdots, y_{n-1}$; let a and b (α and β) be the minimum and maximum of x in $M(\bar{H})$. Then $\alpha \leqq x \leqq \beta$ is both the ordinary and the essential supporting strip of H or $\bar{H}$ normal to the x-axis, but the essential supporting strip of M normal to the x-axis will in general have the form $a' \leq x \leq b'$ with $a \leq a' < b' \leq b$.

For $a \leq d \leq b$ ($\alpha \leq \delta \leq \beta$) denote by $M(d)$ ($H(\delta)$) the intersection of the halfspace $x \leq d$ ($x \leq \delta$) with $M(H)$, and by $m(d)$ ($h(\delta)$) the intersection of $M(H)$ with $x = d$ ($x = \delta$). Then $H(\alpha) = h(\alpha) = h(\beta) = 0$. If $|X|'$ is the $(n-1)$-dimensional measure of a set X in a hyperplane, then

$$(14) \qquad |M(d)| = \int_a^d |m(x)|' dx, \qquad H(\delta) = \int_\alpha^\delta |h(z)|' dz.$$

Following the original idea of Brunn we introduce the function $z(x)$ by the relation

$$|H(z)| = |M(x)|, \qquad a \leq x \leq b.$$

$z(x)$ is a monotone increasing, absolutely continuous function of x with

$$(15) \qquad \begin{cases} z(x) = \alpha & \text{for } a \leq x \leq a' \\ z(x) = \beta & \text{for } b' \leq x \leq b \\ \alpha < z(x) < \beta & \text{for } a' < x < b'. \end{cases}$$

Because of $|h(z)|' > 0$ for $\alpha < z < \beta$ the relations (14) and (15) show that

$$(16) \qquad \begin{cases} \dfrac{dz}{dx} = 0 & \text{for } a \leq x < a' \text{ and } b' < x \leq b \\ \dfrac{dz}{dx} = \dfrac{|m(x)|'}{|h(z(x))|'} & \text{for almost all } x \text{ with } a' \leq x \leq b'. \end{cases}$$

Now construct the set M^* in the strip $a + \rho\alpha \leq d^* \leq b + \rho\beta$ by defining its intersection $m^*(d^*)$ with the plane $x = d^*$ as follows: Because $d + \rho z(d)$ is a strictly monotone and continuous function of d, a given d^* in $[a + \rho\alpha, b + \rho\beta]$ determines exactly one d in $[a, b]$ such that $d + \rho z(d) = d^*$. We then put

$$(17) \qquad \begin{aligned} m^*(d^*) &= m(d) + \rho h(z(d)) \text{ if } a' < d < b' \text{ and } m(d) \neq 0 \\ &= 0 \text{ for all other } d \text{ in } [a, b]. \end{aligned}$$

Since $h(z(d))$ lies in $x = z(d)$ and $m(d)$ in $x = d$, the set $m^*(d^*)$ lies really in $x = d + \rho z(d) = d^*$, and since $\rho h(z(d)) \subset \rho H$ and $m(d) \subset M$ the set M^* is contained in $M + \rho H$.

The inductive hypothesis yields

$$(18) \qquad | m^*(d^*) |' \geq \{ | m(d) |^{1/(n-1)} + \rho \, | h(z(d)) |^{1/(n-1)} \}^{n-1}$$

for the first set of d's in (17). The relations (16) and (18) show that for almost all x in $[a', b']$ with $| m(x) |' > 0$

$$(19) \quad | m^*(x^*) |' \cdot (1 + \rho dz/dx)$$
$$\geq | m(x) |' [1 + \rho \{ | h(z) |' / | m(x) |' \}^{1/(n-1)}]^{n-1} (1 + \rho (| m(x) |' / | h(z) |')).$$

But

$$(20) \qquad (1 + \rho \lambda^{1/(n-1)})^{n-1} (1 + \rho \lambda^{-1}) \geq (1 + \rho)^n \text{ if } \rho > 0, \ \lambda > 0$$

and the equality sign holds only for $\lambda = 1$ (compare [5] or [6, p. 61, Theorem 64]). (19) and (20) yield

$$(21) \qquad | m^*(x^*) |' (1 + \rho (dz/dx)) \geq | m \, x) |' \cdot (1 + \rho)^n$$

for almost all x in $[a', b']$ with $| m(x) |' > 0$. But (21) holds for all x with $| m(x) |' = 0$. The definition of $z(x)$ and (15) show that $| m(x) |' = 0$ a. e. in $[a, a']$ and $[b', b]$, so that (21) holds a. e. in $[a, b]$. Hence

$$(22) \quad | M + \rho H | \geq | M^* | = \int | m^*(x^*) |' \, dx^*$$
$$= \int_a^b | m^*(x^*) |' (1 + \rho (dz/dx)) \, dx$$
$$\geq (1 + \rho)^n \int_a^b | m(x) |' \, dx = (1 + \rho)^n \, | M | = \{ | M |^{1/n} + \rho \, | H |^{1/n} \}^n.$$

This proves the inequalities (10) and (12). We also observe that the equality sign can hold in (12) only if $| M + \rho H | = | M^* |$. This implies that $a' = a$ and $b' = b$, because M^* contains no points in $x < a' + \rho \alpha$ and $M + \rho H$ intersects any strip $a + \rho \alpha < x < c$ in a set of positive measure, because $m(a) \neq 0$. Varying the direction of the x-axis we find

(23) *If the equality sign holds in* (12), *then* $\bar{M}_H = \bar{M}$.

It is now quite simple to discuss the equality sign in (12), but (23) and Theorem 1 contain Theorem 2, because—as pointed out by Dinghas and Schmidt—the equality sign in (12) for some $\rho > 0$ implies equality for any $\rho' < \rho$. For if $\rho = \rho' + \delta$ then $M + \rho H = (M + \rho' H) + \delta H$ hence

$$|M|^{1/n}+\rho|H|^{1/n}=|M+\rho H|^{1/n}\geq|M+\rho' H|^{1/n}+\delta|H|^{1/n}$$
$$\geq|\bar{M}|^{1/n}+\rho'|H|^{1/n}+\delta|H|^{1/n}=|M|^{1/n}+\rho|H|^{1/n}.$$

Therefore the equality in (12) implies (for $\rho\to 0$) equality in (10).

4. That the equality sign holds in (10) when $\bar{M}_E$ is homothetic to L and $B(M)=B(M_E)$ is trivial. But it is worth noticing that the condition $B(M)=B(M_E)$ is not superfluous. For if N is a sequence of points which tends to ∞ and $M=L\cup N$, then $\bar{M}_E=L$ but $B(M)>B(L)$. Examples of W. H. and G. C. Young in [11, pp. 276-281] show that the same effect may be obtained with suitable convergent sequences.

(8) and the formulation of Theorem 1 show that we may restrict ourselves for the "only if" part in the discussion of the equality sign to the case of *closed* M. We show first:

(24) *If the equality sign holds in* (10) *then* $|m(d)|'>0$ *for* $a'<d<b'$. First let $a<d<b$, put $M(d)=M'$, and call M'' the intersection of M with $x\geq d$. Let m'_ρ and m''_ρ be the intersections of $M'+\rho H$ and $M''+\rho H$ with $x=d$. If P denotes the orthogonal projection of H on $x=0$ then the intersection of $M'+\rho H$ $(M''+\rho H)$ with $x\geq d$ $(x\leq d)$ is contained in the cylinder with base $m'_\rho+\rho P$ $(m''_\rho+\rho P)$ and altitude $(\beta-\alpha)\rho$ (the width of ρH in the direction of the x-axis). Therefore

$$|(M'+\rho H)\cap(M''+\rho H)|\leq(\beta-\alpha)\rho(|m'_\rho+\rho P|'+|m''_\rho+\rho P|').$$

Since $M+\rho H=(M'+\rho H)+(M''+\rho H)$ it follows that

$$|M+\rho H|\geq|M'+\rho H|+|M''+\rho H|$$
$$-(\beta-\alpha)\rho(|m'_\rho+\rho P|'+|m''_\rho+\rho P|'),$$

hence because of $|M|=|M'|+|M''|$,

$$\rho^{-1}(|M+\rho H|-|M|)\geq\rho^{-1}(|M'+\rho H|-|M'|)$$
$$+\rho^{-1}(|M''+\rho H|-|M''|)$$
$$-(\beta-\alpha)(|m'_\rho+\rho P|'+|m''_\rho+\rho P|'). \tag{25}$$

Because $m'_\rho\supset m(d)$, $m''_\rho\supset m(d)$, and $m(d)$ is closed, both $m'_\rho+\rho P$ and $m''_\rho+\rho P$ tend for $\rho\to 0$ to $m(d)$, and it follows from (7) and (25) that

$$B(M)\geq B(M')+B(M'')-2(\beta-\alpha)|m(d)|'. \tag{26}$$

Assume now that $a'<d<b'$. Then $|M'|>0$, $|M''|>0$ and (10) and (26) yield

$$
\begin{aligned}
B(M) &\geq n\,|L|^{1/n}(|M'|^{(n-1)/n} + |M''|^{(n-1)/n}) - 2(\beta - \alpha)\,|m(d)|' \\
&= n\,|M|^{(n-1)/n}\,|L|^{1/n}(|M'|^{(n-1)/n}\,|M|^{(1-n)/n} \\
&\qquad + |M''|^{(n-1)/n}\,|M|^{(1-n)/n}) - (\beta - \alpha)\,|m(d)|'.
\end{aligned}
$$

Because of $|M'|\cdot|M|^{-1} + |M''|\cdot|M|^{-1} = 1$

$$|M'|^{(n-1)/n}\,|M|^{(1-n)/n} + |M''|^{(n-1)/n}\,|M|^{(1-n)/n} > 1,$$

so that

$$B(M) > n\,|M|^{(n-1)/n}\,|L|^{1/n} - (\beta - \alpha)\,|m(d)|'.$$

Hence if $m(d) = 0$ the equality sign cannot hold in (10), which proves (24).

Putting $\lambda = \lambda(x) = |h(z(x))|'/|m(x)|'$ it follows from (24) and (19) that equality in (10) implies that a. e. in $[a', b']$

$$|m^*(x^*)|'(1 + \rho(dz/dx)) \geq |m(x)|'\{1 + \rho[(n-1)\lambda^{1/(n-1)} + \lambda^{-1}]\};$$

hence (compare (22))

$$
\begin{aligned}
|M + \rho H| \geq |M^*| &\geq \int_{a'}^{b'} |m(x)|\,dx \\
&\quad + \rho \int_{a'}^{b'} |m(x)|'[(\mathrm{n}-1)\lambda^{1/(n-1)} + \lambda^{-1}]dx
\end{aligned}
$$

and

$$\rho^{-1}(|M + \rho H| - |M|) \geq \int_{a'}^{b'} |m(x)|'[(n-1)\lambda^{1/(n-1)} + \lambda^{-1}]dx.$$

Now

$$(n-1)\lambda^{1/(n-1)}(x) + \lambda^{-1}(x) \geq n$$

and the equality sign holds only for $\lambda(x) = 1$. This can be seen by differentiation or from the theorem of the arithmetic and geometric mean (see [6, p. 17]).

Therefore (since we assumed $|M| = |L|$)

$$B(M) > n\,|M| = n\,|M|^{(n-1)/n}\,|L|^{1/n}$$

unless $\lambda(x) \neq 1$ a. e. in $[a', b']$.

The absolute continuity of $z(x)$, the equality in (10), and (16) imply therefore that $z(x) = x + \alpha - a'$ for $a' \leq x \leq b'$. If we place L so that its center of gravity coincides with that M, then $\lambda(x) = |h(z)|'/|m(x)|' = 1$ a. e. shows that the essential supporting strip $a' \leq x \leq b'$ of M must be a supporting strip of L.

As a convex body, L is the intersection of its supporting strips. Varying the direction of the x-axis we see that $M_E \subset L$. If a point $p \in L - M_E$

existed, then a sphere S with center p and disjoint from M_E would exist because M_E is closed. Then $|S \cap L| > 0$ because L is convex, hence $|L| > |M_E|$, which contradicts (9) and $|L| = |M|$.

This proves that L and M_E are homothetic and

$$B(M_E) = n\,|\bar{M}|^{(n-1)/n}\,|L|^{1/n},$$

if the equality sign holds in (10). Therefore also $B(M) = B(M_E)$.

5. The statement (11) solves the main problem for convex C and those D for which $\int_D \sigma(u)\,dS \geq B(M)$. Unfortunately the class of surfaces for which this relation holds, even when $\sigma(u) \equiv 1$, is very restricted. Since the best possible results can doubtless not be obtained through using $B(M)$, we mention only one class Γ' for which the relation

$$\int_D \sigma(u)\,dS = B(M) \tag{27}$$

is very easily established: Γ' consists of topological spheres of class D' where in addition the set of points in which the first partial derivatives are continuous can be covered by a finite number of neighborhoods in which these partials satisfy a Lipschitz condition. The class Γ is defined to consist of Γ' and all closed convex surfaces.

(28) *If C is convex then the surfaces homothetic to K and only these minimize $\int_D \sigma(u)\,dS$ among all surfaces in Γ with a given volume.*

From this theorem we derive the corresponding theorem for *arbitrary C*. Let $\bar{C}$ denote the boundary of the convex closure of C and let $\bar{\sigma}^{-1}(u)\cdot u$ be the point of the form $\lambda\cdot u$, $\lambda > 0$, on $\bar{C}$. Then $\sigma(\bar{u}) \leq \sigma(u)$ hence

$$\int_D \bar{\sigma}(u)\,dS \leq \int_D \sigma(u)\,dS \tag{29}$$

for every surface D in Γ. We are going to show that the equality sign holds in (29) for surfaces D which are homothetic to the polar reciprocal K of $\bar{C}$ with respect to $|x| = 1$. It suffices, of course, to prove this for K itself.

Every point p of $\bar{C}$ lies on an $(n-1)$-simplex with vertices in C (see [1, p. 9]), and therefore on a simplex $T(p)$ of lowest dimension $\nu \leq n-1$ with vertices on C. Then $T(p) \subset \bar{C}$. This is obvious for $\nu = 0$. For $\nu > 0$ the point p is an interior point of T (considered as set in a ν-dimensional plane), and a hyperplane through p either contains $T(p)$ or separates at least two of the vertices of $T(p)$. A supporting plane π of $\bar{C}$ through p must therefore contain T, and since $\pi \cap \bar{C}$ is convex, $T(p) \subset \bar{C}$.

If $\nu \geq 1$ and $q_0, q_1, \cdots, q_\nu$ are the vertices of $T(p)$, then p has the form

$$p = \Sigma\, \epsilon_i q_i, \qquad \epsilon_i > 0, \qquad \Sigma\, \epsilon_i = 1, \tag{30}$$

hence, after defining $\bar{\sigma}(x)$ in the same way as $\sigma(x)$ for $|x| \neq 1$,

$$\begin{gathered} 1 = \bar{\sigma}(p) = \bar{\sigma}(\Sigma\, \epsilon_i q_i) \leq \Sigma\, \epsilon_i \bar{\sigma}(q_i) = 1 \text{ or} \\ \bar{\sigma}(p) = \Sigma\, \epsilon_i \bar{\sigma}(q_i) \end{gathered} \tag{31}$$

If $\bar{\sigma}(x)$ is now interpreted as supporting function of a body L with boundary K then (30) and (31) imply that at a point of K where the exterior normal is parallel to p, more than one, namely $\nu + 1$ linearly independent, supporting planes exist, so that K is not differentiable.

The crucial point is now that on the one hand for any point p of $\bar{C}$, but not on C, the dimension of $T(p)$ is obviously at least one, and on the other hand the points on $\bar{C}$ and not on C are exactly the points $p = \bar{\sigma}^{-1}(u)u$ for which $\bar{\sigma}(u) < \sigma(u)$. Therefore the preceding discussion shows that $\bar{\sigma}(u) = \sigma(u)$ at every point of K where a tangent plane exists, so that

$$\int_K \bar{\sigma}(u)\, dS = \int_K \sigma(u)\, dS. \tag{32}$$

If $V(D)$ denotes the volume bounded by the surface D in Γ, then Theorem 1, (27), (28), (29) and (32) contain the following main theorem:

THEOREM 3. *If $\sigma(u)$ is defined for $|u| = 1$, positive and continuous, then for any surface D in Γ*

$$\int_D \sigma(u)\, dS \geq n V^{(n-1)/n}(D)\, V^{1/n}(K)$$

where K is the polar reciprocal to the boundary $\bar{C}$ of the convex closure of C: $\sigma^{-1}(u)u$ with respect to the unit sphere. The equality holds only for D homothetic to K.

Although this theorem shows that the question whether C is convex or not is not of great importance as far as the isoperimetric problem for the Minkowski area $\int \sigma(u)\, dS$ is concerned, there are profound differences regarding other properties of this area. This is made clear by the following theorem which is due to Minkowski (see [9, § 27]):

(33) *The function $\sigma(x)$ (or the surface C) is convex if and only if the following holds: whenever the convex body L_1 contains the convex body L_2 then $\int_{K_1} \sigma(u)\,dS \geq \int_{K_2} \sigma(u)\,dS$, where K_i is boundary of L_i.*

An equivalent condition, which will be discussed elsewhere, is that the hyperplanes be minimal (that is area minimizing) surfaces.

6. Theorem 3 leads also to a solution of the *isoperimetric problem in Minkowski geometry.* Let $x_1, x_2, \cdots, x_n$ be a rectangular coordinate system in a Euclidean space associated with a given Minkowski metric μ (for this and the following see section 2 in [2]). Then a positive homogeneous function $f(x)$ of degree 1 exists, which is positive for $x \neq 0$ and such that $\mu(x, y) = f(x - y)$. Area and volume are defined only if μ is symmetric, that is $\mu(x, y) = \mu(y, x)$ or $f(x) = f(-x)$. The surface T: $f(x) = 1$ is the Minkowskian unit sphere. It bounds the closed convex set $S: f(x) \leq 1$.

If $| X |$ denotes the exterior Lebesgue measure of the set X with respect to the associated Euclidean space, then X has the *Minkowski measure*

$$| X |_m = \sigma | X | \tag{34}$$

where

$$\sigma = \omega^{(n)} | S |^{-1} \text{ and } \omega^{(n)} = \pi^{n/2}\Gamma^{-1}(n + 2^{-1})\,;$$

that is, $\omega^{(n)}$ is the volume of the Euclidean unit sphere.

The intrinsic area of a surface in a Minkowski space is defined in terms of the one-dimensional metric μ, but admits for hypersurfaces D of class D' an integral representation of the form $\int \sigma(u)\,dS$, which is obtained as follows: Let the tangent hyperplane of D at a point where it exists, have normal u (because of the symmetry of μ or S we need not distinguish between interior and exterior normals). The hyperplane normal to u through the origin intersects S in an $(n-1)$-dimensional convex set $S(u)$. If $| X |'$ denotes exterior $(n-1)$-dimensional Lebesgue measure with respect to the associated Euclidean space, then

$$\sigma(u) = \omega^{(n-1)} / | S(u) |'.$$

The intrinsic area of a topological sphere D of class D' is then

$$A_m(D) = \int_D \sigma(u)\,dS \tag{35}$$

where dS is, of course, the Euclidean surface element (see the next section).

Since convex surfaces are in general not of class D' this representation has not yet been established for the whole class Γ, but this gap will be filled in the next section.

The surface $C:\sigma^{-1}(u)u$ *is convex.* (A proof is found in the author's paper "A theorem on convex bodies of the Brunn-Minkowski type," *Proceedings of the National Academy of Sciences,* vol. 35 (1949), pp. 27-31). If K denotes the polar reciprocal of C with respect to $|x|=1$ and $V_m(D)$ is the Minkowski volume (34) enclosed by the surface D, then Theorem 3, (34) and (35) yield

$$A_m(D) \geqq nV_m^{(n-1)/n}(D)\,V_m^{1/n}(K)\sigma^{-1}. \tag{36}$$

The factor σ^{-1} appears because K depends on the choice of the coordinate system x. But "homothetic" is independent of the coordinate system and has a Minkowski meaning. K can be characterized in terms of perpendicularity in a very similar manner as in the two dimensional case (see [3, p. 867]), but this will be discussed elsewhere.

For a satisfactory Minkowski isoperimetric inequality it is desirable to eliminate the factor σ^{-1} which depends on the associated Euclidean metric. This can be most easily accomplished by choosing among all the surfaces homothetic to K the surface $K^* = \sigma^{-1}K$. Then

$$V_m^{1/n}(K^*) = V_m^{1/n}(K)\sigma^{-1}. \tag{37}$$

A Minkowskian characterization of K^* is obtained by substituting K^* for D in (36):

$$A_m(K^*) = nV_m(K^*). \tag{38}$$

This relation may be considered as the Minkowskian analogue to the Euclidean fact that $n\omega^{(n)}$ is the surface of the unit sphere. (The relation $A_m(T) = nV_m(T)$ is in general not correct.) Since the associated Euclidean space to a given Minkowski metric is determined up to a non-degenerate affine transformation, *the class* Γ *is Minkowski invariant.* Therefore we may formulate the solution of the *Minkowski isoperimetric problem* as follows:

THEOREM 4. *For all surfaces* D *in* Γ

$$A_m(D) \geqq nV_m^{(n-1)/n}(D)\,V_m^{1/n}(K^*),$$

where K^* *is homothetic to the polar reciprocal of* C *with respect to* $|x| = 1$ *and* $A_m(K^*) = nV_m(K^*)$. *The equality sign holds if and only if* D *is homothetic to* K^*.

The choice of K^* has the great advantage that it permits us to generalize the fundamental relation (7) to Minkowskian·geometry. It is not true that the intrinsic area of D is the limit of $\rho^{-1}(|M + \rho S|_m - |M|_m)$, where M is the interior of D. But

THEOREM 5. *If* H^* *is the interior of* K^*, *then for any surface* D *in* Γ *with interior* M

$$A_m(D) = \lim \rho^{-1}(|M + \rho H^*|_m - |\bar{M}|_m).$$

Proof. For the interior H of K the relations (27), (34) and (35) yield

$$\int_D \sigma(u)\,dS = A_m(D) = \lim \rho^{-1}(|M + \rho H| - |\bar{M}|)$$
$$= \lim (\sigma\rho)^{-1}(|M + \rho H|_m - |\bar{M}|_m)$$
$$= \lim (\sigma\rho)^{-1}(|M + \rho\sigma(\sigma^{-1}H)|_m - |\bar{M}|_m)$$

which proves the Theorem because $\sigma^{-1}H = H^*$.

In analogy to (7) we may now define as Minkowski area $B_m(X)$ of the boundary of a set X in a Minkowski space the number

$$B_m(X) = \liminf_{\rho\to 0+} \rho^{-1}(|X + \rho H^*|_m - |\bar{X}|_m), \tag{39}$$

and Theorem 1 yields

THEOREM 6. *For any set* M *with* $0 < |\bar{M}|_m < \infty$

$$B_m(M) \geq n\,|\bar{M}|_m^{(n-1)/n}\,|H^*|_m,$$

and the equality sign holds if and only if $\bar{M}_E$ *is homothetic to* $\bar{H}^*$ *and* $B_m(M) = B_m(M_E)$.

7. To make the preceding result valid, (35) must be established for convex surfaces. The results and notations of [2] will be used freely. Since convex surfaces in a Minkowski space are birectifiable, (35) is contained in the following general theorem:

THEOREM 7. *If* (P, x) *is a* k*-dimensional rectifiable manifold in a Finsler space* F *with local rectifiable representation* $x(u)$, *then*

$$(40) \qquad | P, x | = | P, x |_k{}^H = \int_p \sigma(u) \; {}^{(u,x)}\Delta^{(u,x)}(u)\, du_1 \cdot \cdot \cdot du_k$$

where $\sigma^{(u,x)}$ is the term $\omega^{(k)}\Xi^{(u,x)}$ of [2, (7, 6)] (the k's in $\Xi^{(u,k)}$ and $\Delta^{(u,k)}$ are misprints) and means the following:

The derivatives $\partial x_i/\partial u_j$ exist for a definite u-system and a definite x-system almost everywhere (a. e.) in the u-coordinate neighborhood. If the matrix $\partial x/\partial u = \partial x_i/\partial u_j) \;{}_{\substack{i=1,\ldots,n \\ j=1,\ldots,k}}$ has rank k then the k-dimensional plane in the tangent space ξ at $x(u)$ spanned by the k vectors $\partial x/\partial u_j$ intersects the set $\Phi(x, \xi) = 1$ in a k-dimensional convex set whose k-dimensional measure with respect to ξ considered as a rectangular coordinate system is $1/\Xi^{(u,x)}(u) = \omega^{(k)}/\sigma^{(u,x)}(u)$. Since $\Delta^{(u,k)}(u)\, du_1 \cdot \cdot \cdot du_k$ is the Euclidean surface element of (P, x) at $x(u)$ (see [2, (7.5)]) (40) has the form (35) (but u does not have the same meaning), moreover $\Delta^{(u,k)}(u) = 0$ if $\partial x/\partial u$ has smaller rank than k. Therefore $\sigma^{(u,x)}(u)$ need not be defined in that case, we simply put the whole integrand $\sigma^{(u,x)}(u)\Delta^{(u,k)}(u) = 0$.

Because of the additivity properties of $| P, x |$, $| P, x |_k{}^H$, and the integral, the general case can be reduced to the local case where P is a bounded closed convex set with positive measure $| P |'$ in E^k with rectangular coordinates $u = (u_1, \cdot \cdot \cdot, u_k)$ and $x(u)$ moves in one coordinate neighborhood (x) of P. We choose these coordinates once and for all, and can then replace $\sigma^{(u,x)}(u)$ and $\Delta^{(u,x)}(u)$ by $\sigma(u)$ and $\Delta(u)$. We show first that

$$(41) \qquad | P, x |_k{}^H = \int_P \sigma(u)\Delta(u)\, du.$$

The paper [2] contains a mistake regarding the measure $| P, x |_k{}^H$. The relation $| P_1, \pi_1{}^\epsilon |_n = \sum_K | x(P_1 V_k), \delta |_n$ on p. 254 is in general not correct. Therefore the proof that $| P, x |_n$ is monotone does not hold, and, in fact, this measure is not monotone. This follows from another representation for $| P, x |_n{}^H$ which Prof. H. Federer communicated to the author. If P is the union of a countable number of compact sets in some metric space, $x(p)$ maps P continuously into a metric space with distance δ, and $\nu(x_0)$ denotes the number of (maximal connected) components of the set of points p in P with $x(p) = x_0$, then

$$(42) \qquad | P, x |_n{}^H = \int_{x(P)} \nu(x)\, d\, | M, \delta |_n,$$

where the right side means the integral of $\nu(x)$ over $x(P)$ with respect to

n-dimensional Hausdorff measure on $x(P)$. It follows from (42) that $|P, x|_n^H$ is invariant under the change of the parametrization, that is [2, (4.8)] (the proof given in [2] does not hold because it uses monotoneity.) It is easy to see that the integral in (42) is not monotone.

It is known that (41) holds if it has been established for an F_σ subset of P, on which the mapping $p \to x(p)$ is one-to-one, compare H. Federer and A. P. Morse, "Some properties of measurable functions," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 270-277, Theorem 3.1 and Section 5. This reference implies the mentioned reduction for the integral over the number $N(x_0)$ of points p with $x(p) = x_0$, but because of the rectifiability $N(x) \neq \nu(x)$ on a set M_0 with $|M_0, \delta|_n = 0$.

We therefore consider an F_σ subset $\bar{P}$ of P on which $p \to x(p)$ is one-to-one. It follows from (42) that $|\bar{P}, x|_k^H = |x(\bar{P}), \delta|_k$. Rectifiability means the existence of a constant β such that the distance δ in F satisfies the relation

$$\delta(x(u), x(v)) \leq \beta\, |u - v| \quad \text{for } u, v \text{ in } P. \tag{43}$$

The partials $\partial x_i/\partial u_j$ exist therefore a. e. in P and are bounded. Consequently $\Delta(u)$ exists a. e. and is bounded. $\sigma(u)$ is bounded because P is closed.

For a given $\epsilon > 0$ there exists a compact subset P' of $\bar{P}$ with $|\bar{P} - P' < \epsilon$ such that $x(u)$ is uniformly totally differentiable on P'. By Kolmogoroff's Principle $|\bar{P} - P', x|_k^H \leq \beta^k |\bar{P} - P'| \beta^k \epsilon$ hence

$$\begin{aligned} |x(P'), \delta|_k \leq |x(\bar{P}), \delta|_k = |\bar{P}, x|_k^H < |P', x|_k^H \\ + |\bar{P} - P', x|_k^H \leq |x(P'), \delta|_k + \beta^k \epsilon. \end{aligned} \tag{44}$$

The boundedness of $\Delta(u)\sigma(u)$ and (44) show that it suffices to establish (41) for P'.

The uniform total differentiability of $x(u)$ on P' implies: for every point $u_0 \in P'$ there is a $\rho_1 > 0$ such that for $u_i \in P' \cap S(u_0, \rho_1)$, $i = 1, 2$, ($S(u_0, \rho_1)$ is the set of all points u in E^k with $|u - u_0| < \rho_1$).

$$x(u_1) - x(u_2) = y(u_1) - y(u_2) + z\, |u_1 - u_2| \text{ with } |z| < \epsilon.^4 \tag{45}$$

We put

$$y(u) = x(u_0) + (u - u_0)\, \partial x(u)/\partial u. \tag{46}$$

If $\Delta(u_0) \neq 0$, then

---

[4] (45) is stated in [7, p. 363] and proved in [10, p. 751].

(47) $$|y(u_1) - y(u_2)| \geq 2k(u_0)|u_1 - u_2|\ ;\ k(u_0) > 0.$$

There also is a $\rho_2 > 0$ such that for $u \,\varepsilon\, P' \cap S(u_0; \rho_2)$

(48) $$\sigma(u_0)\Delta(u_0)/(\sigma(u)\Delta(u)) = 1 + \theta \text{ with } |\theta| < \epsilon.$$

(45) and (47) imply the existence of a $\rho_3 > 0$, $\rho_3 \leq \rho_1$, such that

(49) $$|x(u_1) - x(u_2)| \geq k(u_0)|u_1 - u_2| \text{ for } u_i \,\varepsilon\, P' \cap S(u_0, \rho_3).$$

With the notations

(50) $$w = x(u_1) - x(u_2),\ v = y(u_1) - y(u_2),\ \bar{w} = w/|w|,\ \bar{v} = v/|v|$$

we conclude from

$$|\bar{w} - \bar{v}| = \left|\frac{w(|v| - |w|) + |w|(w - v)}{|v| \cdot |w|}\right| \leq \frac{2|w - v|}{|v|},$$

(45), and (49) that

(51) $$|\bar{w} - \bar{v}| < 2|z|/k \text{ for } u_i \,\varepsilon\, P' \cap S(u_0, \rho_3).$$

We put $\Phi_0(\xi) = \Phi(x(u_0), \xi)$. Then for $|\xi| = 1$ and a suitable $\rho' > 0$

(52) $$\Phi_0(\xi)/\Phi(x, \xi) = 1 + \theta \text{ with } |\theta| \leq \epsilon \text{ when } |x - x(u_0)| < \rho'.$$

Because Φ is homogeneous in ξ the relation (52) holds for all $\xi \neq 0$. If δ_0 is the (Minkowski) distance which corresponds to Φ_0 as integrand, then it follows from (52) that

(53) $$\delta_0(x(u_1), x(u_2))/\delta(x(u_1), x(u_2)) = 1 + \theta \text{ with } |\theta| \leq \epsilon,$$

when $u_i \,\varepsilon\, S(u_0, \rho_4)$ with a suitable $\rho_4 > 0$. Moreover

$$\Phi_0(w)/\Phi_0(v) = 1 + \frac{|w|\Phi_0(\bar{w}) - |v|\Phi_0(\bar{v})}{|v|\Phi_0(\bar{v})}$$
$$= 1 + \frac{\Phi_0(\bar{w}) - \Phi_0(\bar{v})}{\Phi_0(\bar{v})} + \frac{|w| - |v|}{|v|\Phi_0(\bar{v})}\Phi_0(\bar{w}).$$

Since $||w| - |v|| \leq |w - v|$ it follows from (45) and (51) that for a suitable positive $\rho_5 \leq \rho_3$

(54) $$\Phi_0(w)/\Phi_0(v) = 1 + \theta \text{ with } |\theta| \leq \epsilon \text{ for } u_i \,\varepsilon\, P' \cap S(u_0, \rho_5).$$

Let $\rho(u_0) = \min(\rho_2, \rho_4, \rho_5)$, then combination of the results (48), (53) and (54) yields for any measurable set $Q \subset P' \cap S(u_0, \rho(u_0))$

$$(55)\quad \begin{aligned} |x(Q),\delta|_k &= (1+\theta_1)^k |x(Q),\delta_0|_k = (1+\theta_2)^{2k} |y(Q),\delta_0|_k \\ &= (1+\theta_2)^{2k} |y(Q),\eta|_k \sigma(u_0) \\ &= (1+\theta_2)^{2k} \int_Q \Delta(u_0)\sigma(u_0)\,du \\ &= (1+\theta_3)^{2k+1} \int_Q \sigma(u)\Delta(u)\,du, \end{aligned}$$

where $\eta(x,y)$ is the Euclidean distance $|x-y|$ and $|\theta_i| \leq \epsilon$.

Let P'' be the subset of P' where $\Delta(u) \neq 0$. There is a sequence of spheres $S(u^i,\rho(u^i))$ which cover P''. Putting $Q_1 = S_1P''$ and generally $Q_n = P''S_n - S_n \cap (\bigcup_{i=1}^{n-1} Q_i)$ we have $P'' = \bigcup Q_i$ and $Q_i \cap Q_j = 0$. Because the mapping $x(u)$ is one to one on P'', $x(Q_1) \cap x(Q_2) = 0$ and the sets $x(Q_i)$ are measurable, therefore by (55)

$$(56)\quad \begin{aligned} |x(P'),\delta|_k &= \sum |x(Q_i),\delta|_k = \sum (1+\theta^{(i)})^{2k+1} \int_{Q_i} \sigma(u)\Delta(u)\,du \\ &= (1+\theta_4)^{2k+1} \int_{P''} \sigma(u)\Delta(u)\,du. \end{aligned}$$

Since by [7, p. 362] or [10, p. 753]

$$|x(P'-P''),\delta|_k < \beta^k |x(P'-P''),\eta|_k = \beta^k \int_{P'-P''} \Delta(u)\,du = 0,$$

it follows that

$$|P',x|_k^H = \int_{P'} \sigma(u)\Delta(u)\,du.$$

By a method which illustrates convincingly the strength of Kolmogoroff's Principle, Theorem 7 can be reduced to (41) and Section 8 of [2]. As before or in [2, Section 8] the general case can be derived from the special case where P is a bounded closed convex set in E^k and x moves in one coordinate neighborhood in F, so that (P,x) is represented in the form $x(u) = (x_1(u_1,\cdots,u_k),\cdots,x_n(u_1,\cdots,u_k))$ with

$$(57)\qquad \delta(x(u_1),x(u_2)) \leq \beta\,|u_1-u_2| \quad \text{for } u_i \,\varepsilon\, P.$$

As auxiliary space we introduce the E^{n+k} with Cartesian coordinates $x_1,\cdots,x_n,u_1,\cdots,u_k$ and metrize (a portion of) it by the Finsler metric δ^* determined by the integrand

$$(58)\qquad \Phi((x,u),\xi_1,\cdots,\xi_n;\zeta_1,\cdots,\zeta_k) =)\,[\Phi(x,\xi) + |\zeta|]^{\frac{1}{2}}.$$

Then

$$(59)\qquad (x,u)_\epsilon = (x(u),\epsilon u)$$

is a birectifiable surface in the resulting Finsler space because by (57) and (58)

$$\epsilon | u_1 - u_2 | \leq \delta^*(x(u_1), x(u_2)) \leq (\beta^2 + \epsilon^2)^{\frac{1}{2}} | u_1 - u_2 | .$$

The matrix $\partial(x, u)/\partial u$ is

$$(60) \qquad \begin{pmatrix} \partial x_1/\partial u_1 \cdots \partial x_n/\partial u_1 & \epsilon \cdots 0 \\ \cdots \quad \cdots \quad \cdots & \cdots\cdots \\ \partial x_1/\partial u_k \cdots \partial x_n/\partial u_k & 0 \cdots \epsilon \end{pmatrix}.$$

Therefore the expressions $\Delta^\epsilon(u)$ and $\sigma^\epsilon(u)$ corresponding to $\Delta(u)$ and $\sigma(u)$ for the surface (59) have the following properties

$$(61) \qquad \Delta(u) \leq \Delta^\epsilon(u) \leq (\Delta^2(u) + \epsilon^2 B)^{\frac{1}{2}}$$

where B is a constant which depends only on n, k and β (because $| \partial x_i/\partial x_j | \leqq \beta$). For $\epsilon \to 0$ the expression $\Delta^\epsilon(u)$ tends therefore to $\Delta(u)$. The direction coefficients for the tangent space of (59) at a point u are the k-row determinants formed from (60) and divided by $\Delta^\epsilon(u)$. Hence, if $\Delta(u) \neq 0$, the tangent space tends to a k-dimensional space in $(\xi_1, \cdots, \xi_n, 0, \cdots, 0)$ and $\sigma^\epsilon(u) \to \sigma(u)$. Therefore

$$(62) \qquad \sigma^\epsilon(u)\Delta^\epsilon(u) \to \sigma(u)\Delta(u) \text{ when } \epsilon \to 0.$$

By (58) and (59)

$$(63) \qquad \delta^*[(x(u_1), u_1)_\epsilon, (x(u_2), u_2)_\epsilon] \geq \delta^*[x(u_1), u_1)_0, (x(u_2), u_2)_0]$$
$$= \delta^*[(x(u_1), 0), (x(u_2),)].$$

If the set traversed by $(x; u)_\epsilon$, as u traverses P, is denoted by $(x(P), P)_\epsilon$. then by (41) and [2, (8.1)]

$$(64) \qquad | P, (x, u)_\epsilon | = | (x(P), P)_\epsilon, \delta | _k{}^H = \int_p \sigma^\epsilon(u)\Delta^\epsilon(u)\, du$$

and by (41)

$$(65) \qquad | P, (x, u)_0 | _k{}^H = \int_p \sigma^0(u)\Delta^0(u)\, du = \int_p \sigma(u)\Delta(u)\, du = | P, x | _k{}^H.$$

Kolmogoroff's Principle (see [2, p. 248]), (67) and [2, (4.10)] show that

$$(66) \qquad | P, (x, u)_\epsilon | \geq | P, (x, u)_0 | \geq | P, (x, u)_0 | _k{}^H.$$

Theorem 7 is now a consequence of (41), (64), (65) and (66).

UNIVERSITY OF SOUTHERN CALIFORNIA.

REFERENCES.

[1] T. Bonnesen and W. Fenchel, "Theorie der konvexen Körper," *Ergebnisse der Mathematik*, vol. 3, part 1, Berlin, 1934.

[2] H. Busemann, "Intrinsic area," *Annals of Mathematics*, vol. 48 (1947), pp. 234-267.

[3] ———, "The isoperimetric problem in the Minkowski plane," *American Journal of Mathematics*, vol. 69 (1947), pp. 863-871.

[4] ———, "On the problem of Dido," *Studies and Essays Presented to R. Courant on his 60th Birthday*, New York, 1948, pp. 63-73.

[5] A. Dinghas and E. Schmidt, "Einfacher Beweis der isoperimetrischen Eigenschaft der Kugel im n-dimensionalen euklidischen Raum," *Abhandlungen der Preussischen Akademie der Wissenschaften*, 1943, Math. Nat. Klasse no. 7 (1944).

[6] G. H. Hardy, J. E. Littlewood, G. Polya, *Inequalities*, Cambridge, 1934.

[7] A. Kolmogoroff, "Beiträge zur Masstheorie," *Mathematische Annalen*, vol. 107 (1932), pp. 351-366.

[8] L. Lusternik, "Die Brunn-Minkowskische Ungleichnung für beliebige messbare Mengen," *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS* (1935), vol. III (Nouvelle Série), pp. 55-58.

[9] H. Minkowski, "Theorie der konvexen Köper, insbesondre, Begründung ihres Oberflächenbegriffs," *Gesammelte Abhandlungen*, vol. II, Leipzig (1911), pp. 131-229.

[10] G. Nöbeling, "über den Flächeninhalt dehnungsbeschränkter Flächen," *Mathematische Zeitschrift*, vol. 48 (1943), pp. 747-771.

[11] W. H. and G. C. Young, *The Theory of Point Sets*, Cambridge, 1906.

APPROXIMATION IN, AND REPRESENTATION OF, CERTAIN BANACH ALGEBRAS.*

By Richard Arens.

---

1. Introduction. The primary purpose of this paper is to extend to non-commutative Banach algebras the functional representation theorems obtained for certain commutative Banach algebras by M. H. Stone, I. Gelfand, I. Gelfand and M. Neumark, and by I. Kaplansky and the writer (See the bibliography at the end of this paper. We refer to papers listed there by placing the author's name in brackets).

The algebras we study, and call *BQ^*-algebras,* are defined in section **5**. The possibility of treating them by "commutative" methods is due to the fact that we have required each f to commute with its f^* and ff^* to lie in the center (*cf.* 5.02). Our main representation for an arbitrary BQ^*-alegbra A is essentially this: There exists a compact Hausdorff space X on which the group Γ of automorphisms of the quaternions (the orthogonal group) is represented as a transformation-group, and A is isomorphic to the ring of all continuous quaternion-valued functions on X satisfying the conditions of covariance

$$f(\alpha x) = \alpha(f(x)) \qquad \text{(all } x \text{ in } X, \text{ all } \alpha \text{ in } \Gamma). \tag{1.1}$$

(For details see 5.4 and 5.5.) It was shown in [Arens and Kaplansky, 9] that a representation without the covariance feature 1.1, for example, as functions on the structure space (of maximal two-sided ideals) is generally impossible even in the commutative case.

The present methods evolve from the fundamental strategy of Stone, the most important changes being due to the indispensability of the covariance condition. From [Arens II] we predict that all irreducible homomorphisms are quaternion-valued; a commutative-sub-algebra argument based on [Arens and Kaplansky] followed by the extension of maximal ideals shows that the representation by functions on the space of irreducible homomorphisms is norm-preserving (and, *a fortiori,* faithful) and that BQ^*-algebras are semi-simple.

The greatest tactical problem was that of showing that we obtain *all*

---

* Received May 11, 1948.

functions satisfying the condition of covariance 1.1. The first half of the paper is devoted to establishing a suitable generalization, including Stone's generalization, of the Weierstrass approximation theorem. Our investigation goes beyond that needed for quaternions, yielding such results as the following (*cf.* 4.2):

> Let X be a topological space. Let M_n be the algebra of all matrices of degree n with complex entries. Let A be a real linear subalgebra of the algebra of all continuous M_n-valued functions on X. Suppose A contains f^* (the hermitian conjugate) and the real part of the trace of f whenever it contains f. *Then a necessary and sufficient condition that we should be able to approximate a function g uniformly on any compact set by elements from A is that we should be able to do so at any pair of points.*

The paper closes with an unrelated item: it is shown that in a Banach $*$-algebra in which $\| ff^* \| = \| f \|^2$ one cannot have $ff^* = -1$. This is related to, and derives its interest from, an unsolved problem of Gelfand and Neumark (*cf.* 7).

2. Approximation theorems for lattice-valued functions. The principal theorems in this and the following sections are of the following character: there are given two sets of functions A and B with a common domain X and a common range G; moreover,

2.01. *for each x and y in X and any f in B there is a g in A such that*

$$g(x) = f(x) \text{ and } g(y) = f(y).$$

To this we add algebraic and topologic conditions on X, G, and A which insure that A includes B.

We abbreviate condition 2.01 by saying that A is *bi-approximate* to B.

In this section G will be an abelian topological group which is also partially ordered and has defined in it the two lattice-operations satisfying the usual laws. The algebraic structure is related to the order by the familiar condition that $a \leq b$ implies $c - b \leq c - a$ for all a, b, c. Finally it will be supposed that there is a family of symmetric open intervals, containing the group zero element 0, which forms a basis for the topology of G at 0. Consequently, the lattice operations are continuous in both arguments at once. A group G satisfying all these conditions will be called a *topological lattice-ordered abelian group* [*cf.* Birkhoff].

Now a word on the topologization of function spaces. Let G be as above, let X be a topological space, and let $C(X, G)$ be the family of continuous functions on X with values in G. All the operation of G and the ordering of G can be adopted, in a natural way, in $C(X, G)$. But the topology we shall introduce into $C(X, G)$ will not be in terms of the ordering and intervals in $C(X, G)$ (except when X is compact), but will be defined as follows: for each compact subset K of X and each symmetric neighborhood $-e < q < e$ of 0 in G we let (K/e) denote the class of pairs (f, g) of functions in $C(X, G)$ for which $-e < f(x) - g(x) < e$ whenever $x \varepsilon K$. This gives a uniform structure and, *a fortiori*, a topology (called the *k-topology*, see [Arens I]) to $C(X, G)$. It is known that $C(X, G)$ is complete in this uniform structure, if G is complete [Arens I].

The following theorem is an exploitation of a lattice-technical device used by Kakutani [Kakutani, pp. 1004-5] for a formerly similar purpose. It was pointed out to us by Professor Stone (at Harvard, 1942) that this device could be used to prove Stone's theorem (see 2. 3 below) in more or less the way presented here now.

2. 1. THEOREM. *Let G be a*

2. 11 *topological lattice-ordered abelian group.*

Let X be a topological space, and let $C(X, G)$ have the k-topology. Let A be a topologically closed subgroup and

2. 12 *sublattice of $C(X, G)$.*

Then a necessary and sufficient condition that A include a given subset B of $C(X, G)$ is that A be bi-approximate (2. 01) *to B.*

Proof. If A includes B, then obviously, 2. 01 holds. We turn therefore at once to the sufficiency.

Let $f \varepsilon B$, let the compact subset K of X, and let the neighborhood $(-e, e)$ of 0 in G, be given. By 2. 01, for each x, $y \varepsilon K$ we can find $g_{xy} \varepsilon A$ such that $g_{xy}(x) = f(x)$, $g_{xy}(y) = f(y)$. There exists a neighborhood $V(x)$ of x such that $g_{xy}(z) \geqq f(z) - e$ for $z \varepsilon V(x)$. There exist $x_1, \cdots, x_m$ such that K is covered by $V(x_1), \cdots, V(x_m)$. With the corresponding functions, form the least upper bound $g_y = g_{x_1y} \vee \cdots \vee g_{x_my}$. From the properties of this lattice operation we have $g_y(y) = f(y)$, and $g_y(z) \geqq f(z) - e$ for all $z \varepsilon K$. Now we can find a neighborhood $W(y)$ such that $g_y(z) \leqq f(z) + e$ for $z \varepsilon W(y)$. Dualizing the previous compactness argument, we select $y_1, \cdots, y_n$ such that K is covered by $W(y_1), \cdots, W(y_n)$

and then form the greatest lower bound $g = g_{y_1} \wedge \cdots \wedge g_{y_n}$ of the corresponding functions.

It is not hard to see that $-e \leq g(z) - f(z) \leq e$ for all $z \varepsilon K$, or $g - f \varepsilon (K/e)$. But this means that f lies in the closure of A. Since A is closed, $f \varepsilon A$. Hence B is contained in A, as was to be shown.

In the following, R will always denote the real number system, and K the complex system.

2. 2. THEOREM. *Let X be a topological space. Let A be a closed linear subalgebra of $C(X, R)$ with the k-topology. Then a necessary and sufficient condition that A include a given subset B of $C(X, R)$ is that A be bi-approximate to B.*

Proof. We need only show 2. 11 and 2. 12. For 2. 11, we define, as usual, $a \vee b$ to be the larger of the two numbers a and b, etc. To show that the lattice operations are limits of polynomials in A, and thus are performable in A, we use a device suggested by Lebesgue's proof of Weierstrass' theorem [Lebesgue].

If $(1 - (1 - t^2))^{1/2}$ be developed in powers of $1 - t^2$, it converges uniformly to $|t|$ for $|t| \leqq 1$. Now let M be a compact subset of X, let e be a positive real number, and let $f \varepsilon A$ be given. It suffices to consider merely the case in which $|f(x)| \leqq 1$ for $x \varepsilon M$. Select a partial sum $S_n(t)$ of the series mentioned above such that $|S_n(t) - t| < e$ for all t of absolute value not exceeding 1. Then $S_n(f) - S_n(0)$ approximates $|f|$ uniformly to within $2e$ on M. Now $S_n(f) - S_n(0)$ belongs to A because when expanded there is no constant term. Since A is supposed to be closed in the k-topology, the function $|f|$, with values $|f(x)|$, thus approached on M, belongs to A. From here it is only a step to the maximum- and minimum-operations: We set

$$f \vee g, f \wedge g = 1/2(f + g + |f - g|), 1/2(f + g - |f - g|).$$

This proves 2. 2.

2. 3. THEOREM [Stone, pp. 466-469]. *Let X be a compact Hausdorff space. A necessary and sufficient condition that a linear subalgebra A of $C(X, R)$ coincide with $C(X, R)$ itself is that* (a) *A contains the constant functions,* (b) *given $x \neq y$ in X there is an element f in A for which $f(x) \neq f(y)$ and* (c), *A is complete.*

Proof. The necessity is well known; it follows from Urysohn's Lemma. For the sufficiency, we must show that (a) and (b) imply that A is bi-approximate in $C(X, R)$. Let $g \varepsilon C(X, R)$, and let $x, y \varepsilon X$. There is an

$f \varepsilon A$ such that $f(x) \neq f(y)$. By multiplication and addition of suitable constants, we can obtain from f an $h \varepsilon A$ such that $h(x) = g(x)$, $h(y) = g(y)$, as desired. Now (c) shows that A is closed in $C(X, R)$. Hence we apply 2. 2 and obtain 2. 3.

A consequence which is closely related to Weierstrass' theorem itself; and which will be used often in the sequel, will next be presented.

2. 4. LEMMA. *Let X be a topological space. Let $f_1, \cdots, f_n$ be members of $C(X, R)$. Let F be a continuous function of n real variables, and let $F(0, \cdots, 0) = 0$. Then the function $g = F(f_1, \cdots, f_n)$ is in the closed linear subalgebra A generated by $f_1, \cdots, f_n$ in $C(X, R)$ under the k-topology. Moreover, A consists of precisely all such g's.*

Proof. We show that A is bi-approximate to g. Let $x \neq y$ be given. If $f_i(x) = f_i(y)$ for all i then $g(x) = g(y)$, and if $f_1(x), \cdots, f_n(x)$ all vanish then $g(x) = 0$ and so $g(x) = f_1(x)$, $g(y) = f_1(y)$; but if some don't vanish, let $f_i(x) \neq 0$: then t can be found such that $g(x) = tf_i(x) = g(y) = tf_i(y)$. On the other hand, if $f_i(x) \neq f_i(y)$ for some i, then the determinant of the system

$$g(x) = sf_i(x) + tf_i^2(x)$$
$$g(y) = sf_i(y) + tf_i^2(y)$$

does not vanish unless $f_i(x)$ or $f_i(y)$ does. If one of them vanishes, let it be the former, $f_i(x) = 0$. If, in addition, $g(x) = 0$, we can find a t such that $g(x) = 0 = tf_i(x)$, $g(y) = tf_i(y)$, since $f_i(y) \neq 0$; but if $g(x) \neq 0$ then there must be some non-zero $f_j(x)$—in which case we can solve

$$g(x) = sf_i(x) + tf_j(x)$$
$$g(y) = sf_i(y) + tf_j(y),$$

since $f_i(y) \neq 0$.

The converse follows essentially from the completeness of $C(X, R)$. (Cf. [Arens I]).

In the case of complex-valued functions, we deduce the following results.

2. 5. THEOREM. *Let X be a topological space. Let A be a closed real* (i. e., *admitting real scalars*) *linear subalgebra of $C(X, K)$, with the k-topology, and let A contain f^** (*the complex-conjugate*) *whenever it contains f. Then a necessary and sufficient condition that A include a given subset B of $C(X, K)$ is that A be bi-approximate to B.*

Proof. Again we show how 2. 11 and 2. 12 can be fulfilled. We partially order K by saying that $r + pi \geqq 0$ (r, p real) if $r \geqq |p|$. If for complex numbers $r + pi$ we define the *complex number* $\langle r + pi \rangle = m + rpi/m$ where m is $\max(|r|, |p|)$; and then define, for complex q_1, q_2

$$q_1 \vee q_2 = 1/2(q_1 + q_2 + \langle q_1 - q_2 \rangle),$$
$$q_1 \wedge q_2 = 1/2(q_1 + q_2 - \langle q_1 - q_2 \rangle)$$

we obtain two operations which turn out to be the lattice operations consistent with this ordering. We leave the verification to the reader (*Cf.* **3**).

We must now show, after these lattice operations are introduced into $C(X, K)$, that A is closed under them. We shall show that if $f \varepsilon A$ and $f(x) = r(x) + ip(x)$ then

$$\langle f \rangle = m + rpi/m$$

where $m(x) = \operatorname{Max}(r(x), p(x))$, belongs to A. First of all, r and ip separately belong to A, for $r = 1/2(f + f^*)$, $ip = 1/2(f - f^*)$. Next, $p^2 = (ip)(ip)^*$ belongs to A. By 2.5, A contains $F(r, p^2)$ where $F(s, t) = \max(|s|, |t|^{1/2})$. Hence the function m belongs to A. For a real positive u, the function $r(u + m)^{-1}$ belongs to A, by 2.5. Hence $rpi(u + m)^{-1} \varepsilon A$. Now

$$rpi/(u + m) - rpi/m \mid \leqq u.$$

Considering that A is closed, we have $rpi/m \varepsilon A$. Thus $\langle f \rangle \varepsilon A$, and A is a sublattice. This proves 2.5.

From this result, a new flock of corollaries could be deduced; of these we mention one, because of its relation to [Dunford and Segal, Theorem 4].

2.6. Corollary. *Let X be a locally compact space. Let A be a closed real linear subalgebra of $C_0(X, K)$, the class of continuous functions vanishing at infinity* (see [Dunford and Segal]), *and suppose A contains f^* whenever it contains f. Then $A = C_0(X, K)$ if and only if for each x, y in x, $x \neq y$ implies that there is an f in A such that $f(x)$ is real and $f(y)$ is not real.*

The proof is essentially the same as that of 2.3, but about four times as long because (a) the hypothesis is applied once as it stands and then again with x and y interchanged, and (b) only real scalars are to be used, although values may be complex.

Corollary 2.6, from which Theorem 4 [Dunford and Segal] may readily be deduced, differs from it inasmuch Dunford and Segal appear to require complex scalar multiplication in A, while in 2.6 this condition is replaced by requiring essentially that we can obtain all complex numbers at each point y of X. In this form, the results, although not the methods, can be carried over to the case of quaternion valued functions, which are treated in the next two sections.

Our final application is the proof of a theorem which the authors of [Arens and Kaplansky] neglected to prove in the paper just cited. It can be proved by purely ideal-theoretic arguments (as indeed the known special cases are proved), but as we need the result in a later section, we take this opportunity to illustrate the method of bi-approximation in such a connection.

2. 7. THEOREM. *Let X be a locally compact Hausdorff space on which is defined an involutory homeomorphism $(^*)$. Let A be that subring of $C(X, K)$ whose elements satisfy the relation $f(x^*) = f(x)^*$, and vanish at infinity. Then each homeomorphism H of A into K is determined by precisely one point x of X and the formula $H(f) = f(x)$, (f in A).*

Proof. Let the maximal ideal $\mathfrak{a}$ be the set of functions in A which are mapped into 0 by H. By introducing a unit into A (if it has none) and applying [Gelfand, Satz 4], we discover that $\mathfrak{a}$ is a closed set of A and thus a closed linear subalgebra. If we consider the point at infinity w to be adjoined to X, then $C(X, K)$ can be regarded as having the k-topology of $C(X + w, K)$. Suppose that for each x in X there is a g in $\mathfrak{a}$ such that $g(x) \neq 0$. If x and y are points of $X + w$ (and we may exclude the case that x or y is w), there will be two elements g_x and g_y such that $g_x(x) \neq 0 \neq g_y(y)$. By letting $h_x = g_x g_x^*$, $h_y = g_y g_y^*$, we obtain $h_x(x) = a' > 0$, $h_y(y) = b' > 0$. By suitable scalar multiplications, we can make $a' = b' = 1$. Let $h_x(y) = a$, $h_y(x) = b$. Unless $ab = 1$, we can find c and d such that $h = ch_x + ah_y$ which has the value 1 at x and y. If $ab = 1$ and $a \neq 1$, we can find c and d such that $h = ch_x + dh_x^2$ has the value 1 at x and y. If $a = 1$, let $h = h_x$. In any event, we have an h in $\mathfrak{a}$ such that $h(x) = 1 = h(y)$. Now let f be any element of A. Then hf belongs to $\mathfrak{a}$, and since $(hf)(x) = f(x)$, $(hf)(y) = f(y)$, we see that $\mathfrak{a}$ is bi-approximate to A. Thus $\mathfrak{a}$ coincides with A. This contradiction shows that there is an x such that $f \varepsilon \mathfrak{a}$ implies $f(x) = 0$. The ideal $\mathfrak{a}$ is clearly not maximal unless there are only two (perhaps coincident) such points, x and x^*, related, as indicated, by the homeomorphism $(^*)$. Since $\mathfrak{a}$ is the kernel of only two possible homomorphisms, which are given by x and x^*, one of these must determine H by the formula stated.

Thus the space X is the space of homomorphisms of A into K.

3. Preliminary remarks on quarternions and partially ordered groups. The methods of the preceding section can be extended to quaternion-valued functions by setting $q \geqq 0$ for a quaternion q if its real part is non-negative and its pure part lies in the first quadrant for some fixed choice of axes in the system of quaternions, Q. This makes Q a lattice, but the performance of lattice operations in $C(X, Q)$ would require not only the use of f and f^* but also the non-invariant f^i, f^j, and f^k, where by f^i for example, we mean the function with values $-if(x)i$.

The ordering we have found useful reduces, in the special case of Q, to saying $q = r + ai + bj + ck \geqq 0$ if and only if $r \geqq (a^2 + b^2 + c^2)^{1/2}$. This ordering is suggested by the "cone of the future" of special relativity, if the real part r is regarded as the "time." So ordered, Q does not form a lattice (although K does: *cf.* proof of 2.5) because the common part of two "cones of the future" is generally not itself such a cone (*cf.* [Clarkson]). Geometrical considerations make this evident.

But a weaker sort of lattice operation is possible, and the proofs of the next section will utilize it. Its exposition is simplified by the introduction of a singulary operation. We list here some formulae and results which show their *abstract* relationship, and which show how far structures in which they can be defined fall short of lattices. For *concrete* definitions, see the next section.

Let G be a partially ordered abelian group admitting the positive (dyadic) rational numbers as operators, with the requirements: $a, a' \geqq b$ implies $c - b \geqq c - a$ and $1/2(a + a') \geqq b$ for any a, a', b, c in G. Suppose there are defined a singulary operation $\langle\ \rangle$ and a binary operation ∇: then the following 3.01-3.14 are meaningful.

3.01. G is a topological group and ∇ is continuous.

3.02. $a \nabla a = a$ (idempotence).

3.03. $a \nabla b \geqq a, b$ (incidence).

3.04. $a, b \leqq c$ implies $a \nabla b \leqq c$.

With 3.03 and 3.04, G would be a lattice.

3.05. $2(a \nabla b) = 2a \nabla 2b$.

3.06. $(a + b) \nabla (a - b) = a + b \nabla (-b)$.

3.07. $\langle a \rangle = a \nabla (-a)$.

3.08. $a \nabla b = 1/2(a + b + \langle a - b \rangle)$

3.10. G is a topological group and $\langle\ \rangle$ is continuous.

3.11. $\langle 0 \rangle = 0$.

3.12. $\langle a \rangle \geqq \pm a$.

3.13. $a, b \geqq 0$ implies $a + b \geqq \langle a - b \rangle$.

3.14. $\langle 2a \rangle = 2\langle a \rangle$.

We now list a number of propositions. We shall require only the last one for later use. The other are possibly of independent interest. The proofs are very short and are all left to the reader.

3.15. Lemma. *In the presence of* 3.07 *and* 3.08, (3.01, 3.02, 3.03) *is equivalent to* (3.10, 3.11, 3.12).

3.16. Lemma. 3.08 *with* 3.14 *implies* 3.07.

3.17. Lemma. 3.05 *with* 3.06 *and* 3.07 *implies* 3.08.

3.18. Lemma. 3.08 *with* 3.13 *implies* 3.04.

3.19. Lemma. 3.04 *with* 3.07 *implies* 3.13.

3.20. Lemma. *In the presence of* 3.07 *and* 3.08, *G is a lattice if and only if* 3.11, 3.12, *and* 3.13 *hold.*

3.21. Lemma. 3.03 *with* 3.04 *implies* 3.02, 3.05 *and* 3.06 (*case of a lattice*).

3.22. Theorem. 3.08 *with* 3.10, 3.11, 3.12, *and* 3.14 *implies* 3.01, 3.02, *and* 3.03.

4. More general approximation theorems. As ranges for the function spaces later to be considered, we introduce the following spaces. The quaternions form a special case.

Let E be a Banach space with norm $\| \cdot \|$. Form the direct product $G = R \times E$ (*Cf.* [Banach, p. 181]), which is also a real Banach space. If r is a real number, and we speak of $r \varepsilon G$, then we mean the element $(r, 0)$. Whenever $q = (r, p) \varepsilon G$ where $r \varepsilon R$, $p \varepsilon E$, we will denote these components as follows: $r = q'$, $p = q''$. We partially order G by writing

4.01. $q \geqq 0$ when $q' \geqq \| q'' \|$; $q \geqq q_1$ if and only if $q - q_1 \geqq 0$.

4.02. Lemma. *G is a partially ordered abelian topological group* (*in the sense of* **3**).

Proof. The only non-trivial part of the theorem is quickly reduced to the question: if (r, p), $(r_1, p_1) \geq 0$, is $(r + r_1, p + p_1) \geq 0$? We have at once $r \geqq \| p \|$, $r_1 \geqq \| p_1 \|$. From this $r + r_1 \geqq \| p + p_1 \|$ follows, by the triangle inequality for the norm in E.

The singularity operation $\langle\ \rangle$ in G shall be defined as follows:

4.03. $\langle q \rangle = (\operatorname{Max}(| q' |, \| q'' \|),\ q'q''/\operatorname{Max}(| q' |, \| q'' \|)),\ q \neq 0;$
$\langle 0 \rangle = \langle 0 \rangle.$

4.04. LEMMA. *Conditions* 3.10, 3.11, 3.12, *and* 3.14 *hold.*

Proof. The continuity (3.10) of $\langle\ \rangle$ is doubtful at most at the origin; but it is clearly present at the origin also, since

$$\| \langle q \rangle'' \| \leqq \| q'' \| .$$

To prove 3.12, first set $q' = r$, $q'' = p$. Abbreviate $\max(| r |, \| p \|)$ by m. Then

$$| r + m | \leqq m + r, \quad | r - m | \leqq m - r;$$

whence

$$\| p \| \cdot | r \pm m | \leqq m(m \pm r)$$

and

$$\| rp/m \pm p \| \leqq m \pm r$$

or

$$(m \pm r, rp/m \pm p) \geq 0 \text{ in } G.$$

This proves $\langle q \rangle = (m, rp/m) \geq \pm q$. The remaining condition 3.14 obviously holds. Thus 4.04 is proved.

4.05. LEMMA. *If one defines*

$$q \nabla q_1 = 1/2(q + q_1 + \langle q - q_1 \rangle)$$

in G, then this operation is continuous, idempotent, and incidental.

This results from 3.22 and 4.04.

After these preparations, we present our most inclusive approximation theorem.

4.1. THEOREM. *Let $G = R \times E$ where E is a Banach space with norm $\| \cdots \|$. Let X be a topological space. Suppose A is a closed linear subspace of $C(X, G)$ with the k-topology such that*

4.11. *If f, g belong to A then $f'g$ belongs to A;*

4.12. *If f belongs to A then* $(\| f'' \|^2, 0)$ *belongs to* A *(this function has the values* $\| f(x)'' \|^2 \varepsilon R$*); similarly* $(f', 0)$ belongs to A;

4.13. *The elements having* $f''=0$ *form a real linear algebra, (where* $(f',0)(g',0)=(f'g',0)$*; then a necessary and sufficient condition that* A *include a given subset* B *of* $C(X,G)$ *is that* A *be bi-approximate* (2.01) *to* B.

Proof. The condition is obviously necessary. To prove the sufficiency, we will show that for any real positive e, for any compact subset M of X, and for any f in B, one can find a g in A such that for any $z \varepsilon M$ one has $g(z) \geqq f(z) - e$ and $g(z)' \leqq f(z)' + 2e$. This yields $\| g(z)'' - f(z)'' \| \leqq g(z)' - f(z)' + e \leqq 3e$, which shows that g approximates f uniformly on M.

In the construction of g we shall use the ∇ operation; hence we must now show that A is closed under this operation. Again, it is enough to show that for $h \varepsilon A$ one has $\langle h \rangle \varepsilon A$. If $h \varepsilon A$ then $(h',0)$ and $(\| h'' \|^2, 0) \varepsilon A$. Now $(\max(|h'|, \| h'' \|), 0)$ is a continuous function of these which vanishes for zero values of the arguments; hence, by 2.4 and 4.13, this element belongs to A. As for the second component, we observe that for any real positive u, $(h'(u+\max(|h'|, \| h'' \|))^{-1}, 0)$ belongs to A, by 2.4 and 4.13 as before. By applying 4.11, and then subtracting off the first component, we arrive at $(0, h'(u+\max(|h'|, \| h'' \|))^{-1}h'')$. By the same argument used in the proof of 2.5, the limit of this as u approaches zero lies in A. Hence A contains

$$\langle h \rangle = (\max(|h'|, \| h'' \|),\ h'h''(\max(|h'|, \| h'' \|))^{-1}).$$

Now we proceed just as in the proof of 2.1, until we have found $g_{x_1y}, \cdots, g_{x_my}$ which, we recall, have the property that $g_{x_ky}(y) = y$; and $g_{x_ky}(z) \geqq f(z) - e$ for each z in M and proper choice of k. We define $g_y = (\cdots(g_{x_1y} \nabla g_{x_2y}) \nabla \cdots) \nabla g_{x_my} \varepsilon A$. Which way we associate here does not matter, but a specific order must be selected since the associative law does not hold. By 3.01 and 3.02 we have $g_y(y) = f(y)$ and $g_y(z) \geqq f(z) - e$ for each z in K. For each y in M we can find a neighborhood $W(y)$ such that $g_y(z)' \leqq f(z)' + e$. By the compactness of K there are $y_1, \cdots, y_n$ such that M is covered by $W(y_1), \cdots, W(y_n)$. Let us denote the corresponding g_y-functions by $g_1, \cdots, g_n$. The desired g is going to be a linear combination of these g_k. This is where the present proof departs from the method of proving 2.1.

To construct the coefficients of the linear combination, let us first abbreviate by setting $r = \min(g'_1, \cdots, g'_n)$, and let

$$s = 1/(e/n + g'_1 - r) + \cdots + 1/(e/n + g'_n - r).$$

Then for each $k = 1, \cdots, n$, the functions

$$c_k = 1/s \cdot 1/(e/n + g'_k - r) - 1/n$$

are continuous functions of $g_1, \cdots, g_n$ and vanish when these do. By 2.4 and assumption 4.13, the elements $(c_1, 0), \cdots, (c_n, 0)$ belong to A, whence also $c_1 g_1 + \cdots + c_n g_n$ belongs to A. Finally, let $g = c_1 g_1 + \cdots + c_n g_n + (g_1 + \cdots + g_n)/n$. It is apparent now that g lies in A, but we could not immediately have inferred it from the formula

$$4.13 \quad g = g_1/(e/n + g'_1 - r) \cdot 1/s + \cdots + g_n/(e/n + g'_n - r) \cdot 1/s,$$

obtained by collecting terms. For each z in M, the coefficients

$$1/(e/n + g'_1(z) - r(z)) \cdot 1/s(z), \cdots, 1/(e/n + g'_n(z) - r(z)) \cdot 1/s(z)$$

are positive real numbers whose sum is 1. Therefore, since $g_1(z), \cdots, g_n(z)$ lie in the convex set of G defined by $q \geqq' f(z) - e$, we conclude $g(z) \geqq f(z) - e$. Continuing, let us set $g'_1(z) = r_1, \cdots, g'_n(z) = r_n$, and let us suppose that $\min(r_1, \cdots, r_n) = r_k$. Then $s(z) > (e/n + r_k - r_k)^{-1} = n/e$, whence $1/s(z) < e/n$. Now

$$g'(z) = 1/s(z) \cdot r_1/(e/n + r_1 - r_k) + \cdots + 1/s(z) \cdot r_n/(e/n + r_n - r_k);$$

and by the formula for $s(z)$,

$$r_k = 1/s(z) \cdot r_k/(e/n + r_1 - r_k) + \cdots + 1/s(z) \cdot r_k/(e/n + r_n - r_k).$$

Therefore

$$\begin{aligned} g'(z) - r_k = 1/s(z) &\cdot (r_1 - r_k)/(e/n + r_1 - r_k) \\ &+ \cdots + 1/s(z) \cdot (r_n - r_k)/(e/n + r_1 - r_k) \end{aligned}$$

whence

$$0 \leqq g'(z) - r_k < e/n \cdot 1 + \cdots + e/n \cdot 1 = e.$$

Since $r_k = \min(g'_1(z), \cdots, g'_n(z)) \leqq f'(z) + e$ we have $g'(z) < f'(z) + 2e$ as desired. This finishes the proof of 4.1.

We now apply this result to rings of matrix-valued functions. If f is such a function, then f^* will be used to denote the function whose value $f^*(x)$ is the hermitian transpose $f(x)^*$ of $f(x)$, and f^* will be called the *adjoint* of f. By $rtr\, f$ we mean the function having as its value one n-th of the real part of the trace of f.

4.2. THEOREM. *Let X be a topological space. Let M_n be the algebra* *of all matrices of degree n with complex entries. Let A be a closed real linear subalgebra of $C(X, M_n)$ with the k-topology, and suppose*

4.21. *A contains f^* whenever it contains f.*

4.22. *A contains* rtr *f whenever it contains f.*

Then a necessary and sufficient condition that A include a given subset B of $C(X, M_n)$ is that A be bi-approximate to B.

Proof. Let E be the class of matrices in M_n whose trace is zero or imaginary. The class R_0 of real scalar matrices is obviously isomorphic, with preservation of linear combinations, to the real number system R:

$$r \leftrightarrow \begin{pmatrix} r & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & r & \cdot & \cdot & \cdot & 0 \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & r \end{pmatrix}.$$

Observing this isomorphism, we can say that M_n is isomorphic to $R \times E$, with preservation of real linear combinations. The usual topology of M_n is reproduced by taking the "norm" in E to be

$$\| p \| = (\Sigma \mid p_{ij} \mid^2)^{1/2} \qquad\qquad p = (p_{ij}) \, \varepsilon \, E.$$

For f in A, we have $f' \, \varepsilon \, A$ since f' is obviously $rtr\, f$. Since A is a subalgebra, $f'g$ is in A if g is. This shows that 4.11 holds. If $f \, \varepsilon \, A$, then $f - f' = f'' \, \varepsilon \, A$; and $\| f'' \|^2 = rtr(f''f''^*) \, \varepsilon \, A$. Thus 4.12 is verified. Finally, 4.13 holds because A is a real linear subalgebra. An appeal to 4.1 now establishes 4.2.

The earlier results on the real (2.2) and complex (2.5) cases are consequences of this theorem, but of course the real case is already involved in the proof through the medium of 2.4. The following statement for the case of the quaternions Q is now possible.

4.3. THEOREM. *Let X be a topological space. Let A be a closed real linear subalgebra of $C(X, Q)$, with the k-topology, and let A contain f^* (the quaternion-conjugate) whenever it contains f. Then a necessary and sufficient condition that A include a given subset B of $C(X, Q)$ is that A be bi-approximate to B.*

Proof. As is well known, the quaternions can be represented by the ring of unitary matrices of the form

$$\begin{pmatrix} a & -b^* \\ b & a^* \end{pmatrix} \qquad\qquad (a, b \text{ complex});$$

and in this representation, the conjugate quaternion is represented by the hermitian transpose. Moreover, $rtr\, f = 1/2(f + f^*)$ is evidently a member of A if f is. Therefore the truth of 4. 3 is reduced to that of 4. 2.

This theorem is used later in the study of certain representations; but the following application has some interest of its own.

4. 4. THEOREM. *Let X be a completely regular space. Let A be a closed real linear subalgebra of $C(X, Q)$ with the k-topology, and let A contain f^* whenever it contains f. Then a necessary and sufficient condition that $A = C(X, Q)$ is that, given x, y in X and a quaternion q, there is an f in A for which $f(x) = 0$ and $f(y) = q$.*

Proof. The necessity is a consequence of the complete regularity of X. Therefore let us show that A is bi-approximate to $C(X, Q)$. Let $g \varepsilon C(X, Q)$, and let $x, y \varepsilon X$. There are f_x and f_y in A such that $f_x(x) = 0$, $f_x(y) = g(y)$, $f_y(x) = g(x)$, $f_y(y) = 0$. Therefore $f = f_x + f_y$ shares its values at x and y with g, as desired.

5. Representation of *BQ**-algebras.

By a *Banach algebra* we mean a normed ring in the sense of [Gelfand], except we do not require complex scalars nor a unit element. By a *BQ*-algebra* we mean a Banach algebra A in which there is defined an operation $(^*)$ which satisfies, for real r, and f, g in A

5. 01. $(rf + g)^* = rf^* + g^*$, $(fg)^* = g^*f^*$, $f^{**} = f$,

5. 02. ff^* lies in the center of A,

5. 03. $\| f \|^2 \leqq \| ff^* + gg^* \|$, if f and g commute; and the norm itself satisfies

5. 04. $\| fg \| \leqq \| f \| \cdot \| g \|$.

If all conditions except 5. 02 are satisfied, we call A a *real Banach*-algebra.* The letter "Q" is used to imply 5. 02, which was suggested by the quaternions, and the word "real" is then omitted.

We prove first a result on commutative Banach*-algebras.

5. 05. LEMMA. *Let B be a commutative real Banach*-algebra. Let $f \varepsilon B$. Then*

5. 06. $\| x(f) \| \leqq \| f \|$ *and* $x(f^*) = x(f)$ *for any homomorphism x of B into the complex numbers;*

5.07. *there is a homomorphism x of B into the complex numbers for which* $|x(f)| = \| f \|$;

5.08. $ff^* = g^4$ *for some* $g = g^* \varepsilon B$, *and* $ff^* - \| ff^* \|$ *has no inverse in* B.

5.09. *The closure of the principal ideal generated by* ff^* *also contains* f.

Proof. This algebra B can be represented by an algebra A as described in 2.7 above, in such manner that

$$f^*(x) = \overline{f(x)} \text{ and } \| f \| = \sup_{x \varepsilon X} | f(x) | ,$$

by [Arens and Kaplansky, Theorem 9.1]. Now 5.06, 5.07, and 5.08 are obvious consequences of that representation.

We now prove 5.09, also supposing that B is represented by an A as described in 2.7. We designate the ideal by J and ff^* by j. Let $x, y \varepsilon X$ (please refer to 2.7). If f vanishes at x or y, then a scalar t can easily be found such that tjf and f agree in value at x as well as y. Otherwise $0 < j(x), j(y)$, and there is an element

$$h = [(j(x) + j(y))j - j^2]/j(x)j(y)$$

in J. Since $h(x) = h(y) = 1$ it is evident that hf and f agree in value at x as well as y. This shows that J is bi-approximate to f, and by 2.5, J contains f.

Let Hom (A, Q) denote the class of homomorphisms of the ring A into Q.

5.1. LEMMA. *Let A be a BQ^*-algebra. If $x \varepsilon$* Hom (A, Q) *then, for* $f \varepsilon A$,

5.11. $|x(f)| \leqq \| f \|$ ($|\,.\,.\,|$ denotes the modulus in Q),

and

5.12. $x(f^*) = x(f)^*$.

Proof. First, we observe that in a BQ^*-algebra, one has $ff^* = f^*f$. In fact, $f(ff^* - f^*f) = f^2f^* - (ff^*)f = f^2f^* - f(ff^*) = 0$, from which one can deduce that $(ff^* - f^*f)^2 = 0$. Replacing f in 5.03 by $ff^* - f^*f$, and g by 0 we see that $ff^* - f^*f = 0$. Second, we apply 5.06 to the commutative subalgebra A_f generated by f and f^*, giving 5.11 and 5.12.

The next proposition is essentially that Hom (A, Q) has sufficiently many members.

5.2. LEMMA. *Let A be a BQ^*-algebra, and let $f \varepsilon A$. Then there is an element x in* Hom (A, Q) *such that* $| x(f) | = \| f \|$.

Proof. We may exclude the case $f = 0$. Case 1: A has a unit element 1. Let $\| f \| = r$. If $r^2 - ff^*$ has no inverse in A, then, since it lies in the center, it lies in some maximal two-sided ideal M of A. We assert that each g which is not in M has an inverse *mod* M. In fact, if $g \varepsilon M$ then $gg^* \varepsilon M$ by 5.09. Now the center of A is, *mod* the maximal ideal M, a field. Since gg^* lies in the center, there is an h such that $gg^*h - 1$ is in M. This g^*h is the desired inverse *mod* M. Evidently A/M is a division algebra which can, by [Arens II], be only Q, or a subfield of Q. Now $x(r^2 - ff^*) = 0$, whence $x(ff^*) = r^2$. Application of 5.11 and 5.12 gives $| x(f) | = \| f \|$. If $r^2 - ff^*$ *has* an inverse in A, then it has an inverse in the closed commutative linear subalgebra a_f generated by f, f^* and 1. But by 5.08, $r^2 - ff^*$ cannot have an inverse.

Case 2: A has no unit. Imbed A isometrically in a ring A_1 which has a unit, after the fashion of [Arens III, Lemma 4]. We do not care whether 5.03 holds in A_1; at least it holds for the subalgebra A. If $r^2 - ff^*$ has no inverse in A_1, we arrive at the desired homomorphism x of A_1, as before. But x is also a homomorphism of A. But if $r^2 - ff^*$ *has* an inverse $s - h$ in A_1, then $(s - h)(r^2 - ff^*) = 1$. Since A does not contain 1, $sr^2 = 1$, and we arrive at an element $p = r^2h$ in A for which $r^2p = ff^* + pff^*$. Applying 5.08 to the commutative closed linear subalgebra A_f generated by f and f^* we obtain that $ff^* = g^4$, where $g = g^* \varepsilon A_f$. Therefore g^2 is in the center of A. Consider the commutative closed linear subalgebra A_g generated by p, p^*, and g^2. Evidently $\| g^2 \| = r$, and so by the representation of [Arens and Kaplansky, *loc. cit.*], there is at least one homomorphism y of A_g for which $y(g^4) = r^2$. Applying y to the equation $r^2p = g^4 + pg^4$ we obtain the contradiction $r^2 = 0$.

Thus the conclusion of 5.2 is reached in all cases.

If $x \varepsilon \operatorname{Hom}(A, Q)$, and α is an automorphism of Q leaving R invariant, then there is a $y \varepsilon \operatorname{Hom}(A, Q)$ such that

5.3. $$y(f) = \alpha(x(f)) \qquad \text{(for all } f \varepsilon A).$$

In this case we write $y = \alpha x$. In this way the group Γ of automorphisms of Q which leave R fixed acts in a natural way as a group of one-to-one transformations on $\operatorname{Hom}(A, Q)$. (The fact that α is always an inner automorphism and that Γ is isomorphic with the orthogonal group is not relevant at this point, but is implicitly taken into account in the proof of 5.4.) If we introduce into $\operatorname{Hom}(A, Q)$ the weak neighborhood topology, then X becomes a Hausdorff space, and each $f \varepsilon A$ may be regarded as a *continuous* function by setting $f(x) = x(f)$. Moreover, X becomes a compact space. (We are not omitting the 0-homomorphism. But some may prefer to omit it, in which case X is locally compact and 0 becomes the point in infinity.) The proof of these facts can be constructed in the same way as the proof of the related case of the unit sphere of the conjugate space of a Banach space [Alaoglu], or alternately, along the lines of [Arens III, p. 278]. With this topology, the automorphisms act *homeomorphically* on X.

5. 4. THEOREM. *Let A be a BQ^*-algebra. Let $X = \mathrm{Hom}(A, Q)$ be given the weak neighborhood topology. Then X is a compact Hausdorff space. For any automorphism α of Q leaving B fixed, set $y = \alpha x$ if 5. 3 holds. Then α is a homeomorphism of X. If A is imbedded in the natural way in $C(X, Q)$, then A is precisely the class of functions, vanishing at infinity, such that*

5. 41.
$$f(\alpha x) = \alpha(f(x))$$

for all α and all x. Moreover,

5. 42.
$$\| f \| = \sup_{x \varepsilon X} | f(x) |$$

and

5. 43.
$$f^*(x) = f(x)^*.$$

Proof. We establish the facts in the reverse order mentioned. First, 5. 43 and 5. 42 are consequences of 5. 11, 5. 12 and 5. 2 collectively. From 5. 42 and the assumed completeness of 4 we infer that A is a closed real linear subalgebra of $C(X, Q)$ with the k-topology. Each f of A satisfies 5. 41 because of the definition of $\alpha x = y$ by 5. 2; and each f of A vanishes at infinity. Let B be the class of functions for which 5. 41 holds, for all α and x, and which vanish at infinity.

We shall show that A is bi-approximate to B. Let x and y be points of X, and let $f \varepsilon B$. Let $f(x) = a$, $f(y) = b$. The image algebra S in Q of A under x determines, and is determined by the subgroup $x\Gamma$ of those α for which $\alpha x = x$. Since presumably $\alpha a = a$ for exactly those α, we infer that $a \varepsilon S$. Hence there is a $g_x \varepsilon A$ such that $g_x(x) = a = f(x)$. If, furthermore, there is an α such that $y = \alpha x$, then also $b = \alpha a$ and $g_x(y) = g_x(\alpha x) = \alpha g_x(x) = \alpha a = b$; and in this case $h = g_x$.

If there is no α such that $y = \alpha x$, obtain first of all a $g_y \varepsilon A$ such that $g_y(y) = b$ by the same argument used to find g_x. But if $y \neq \alpha x$ for any α, the kernel ideals of x and y must be distinct. Being maximal, each must contain an element not in the other; and therefore there exist h_x and h_y such that $h_x(x) \neq 0$, $h_x(y) = 0$, $h_y(y) \neq 0$, $h_y(x) = 0$. By considering h_x and h_y replaced by suitable scalar multiples of $h_x h_x^*$ and $h_y h_y^*$ respectively, we see that one could achieve $h_x(x) = 1$ and $h_y(y) = 1$. In case this paragraph applies, let $h = h_x g_x + h_y g_y$. Then $h(x) = a = f(x)$ and $h(y) = b = f(y)$. Such an h was obtained also in the preceeding paragraph. Thus A is bi-approximate to B, and by 4. 3, A coincides with B as asserted in 5. 4. The remaining assertions of 5. 4 were already considered in the introductory paragraph preceding 5. 4.

The proof of 5. 4 should be compared with that of [Arens and Kaplansky, *loc. cit.*]. *There*, the ring A was imbedded in a larger ring A' admitting as scalars a division algebra (K) which was known in advance to contain all possible residue fields; and then previously established results were applied to obtain a representation of A' as functions on a space X. From the way A was imbedded in A' it was then discovered that A consisted of all functions satisfying a certain condition of covariance (*viz.*, $f(x^*) = f(x)^*$) analogous to 5. 41. The same attack was also made on certain algebraic algebras, in [Arens and Kaplansky]. And finally, it had to be established *a posteriori* that X was the space $\mathrm{Hom}(A, K)$ (in 2. 7, above). But *here* (in the proof of 5. 4), we could not "extend" the coefficient field of A from R to Q. This is not only because the new "scalars" would not commute with the ring elements (pre-factors, post-factors, etc.) but a fundamental and unaesthetic asymmetry arises even when one considers the simple case $A = K$ itself. Formal methods (Cayley-Dickson process [Dickson, p. 15]) lead to non-associative algebras, and therefore make a prediction of residue algebras on the basis of [Arens II] impossible. Therefore we began with the space $\mathrm{Hom}(A, Q)$, or in other words, with the space of irreducible (although not inequivalent) representations as linear operators on 2-dimensional unitary space. Hence the analogue of 2. 7, desirable for completeness' sake, was involved from the start. Spaces of ideals, and debates about their topologies, are avoided. On the other hand, the proof that we obtained *all* the continuous functions satisfying a condition of covariance (5. 41) became enormously complicated, requiring almost all of the machinery of sections **3** and **4**.

A smaller distinction between 5. 4 and [Arens and Kaplansky, Theorem 9. 1]—that (*) forms a homeomorphism of X in the latter but not in the former—should also be observed.

A characterization of BQ^*-algebras can be formulated in the following way.

5. 5. THEOREM. *Let X be a compact Hausdorff space on which the orthogonal group Γ acts as a transformation group [leaving a point x_0 fixed]. Consider the subalgebra A of $C(X, Q)$ of those functions [vanishing at x_0] which satisfy*

5. 51. $$f(\alpha x) = J_\alpha(f(x)). \qquad (\textit{for all } x \in X)$$

where J_α *is the inner automorphism of* Q *associated (in the way discovered by Hamilton) with the rotation* α. *Set* $\| f \| = \sup | f(x) |$, *and* $f^*(x) = f(x)^*$ *for each* $f \varepsilon A$. *Then* A *is a BQ*-algebra, (possessing a unit if and only if* x_0 *is isolated). Then* $X = \mathrm{Hom}(A, Q)$; *and all BQ*-algebras are obtainable in this way.*

Proof. The only thing that is not either obvious or covered by 5. 4 is, that $X = \mathrm{Hom}(A, Q)$. But we leave it to the reader to construct the proof of this, along the lines of the proof of 2. 7. The necessary argument of bi-approximity uses the technique of the demonstration of bi-approximity given in 5. 4.

Concerning the topology of $X = \mathrm{Hom}(A, Q)$, we can make the following statement.

5. 6. THEOREM. *Let* A *be a BQ*-algebra, and let* $X = \mathrm{Hom}(A, Q)$ *be given the topology described above. Then* X *is metrizable if and only if* A *is separable. If* X *is metrizable, the metric can be so chosen that it is invariant under the homeomorphisms induced by* Γ.

The proof of [Gelfand, Satz 12] can be adapted to establish the present theorem. The formula for the metric given there, after suitable change of notation and interpretation for quaternion-valued functions, yields a metric invariant under Γ.

We now present a characterization of the space $C(X, Q)$ itself, supposing X to be compact. This theorem is of interest in itself, but it is not likely that the hypothesis could be fulfilled in a ring actually arising in a situation not involving the quaternions intimately *a priori.*

5. 8. THEOREM. *A necessary and sufficient condition that a real Banach algebra* A *is isomorphic to* $C(Y, Q)$ *with*

$$\| f \| = \sup_{x \varepsilon Y} | f(x) |$$

for some compact Hausdorff space Y *is that* A *be a BQ*-algebra containing a subalgebra isomorphic (with preservation of the *-operation) to the quaternions.*

Proof. The necessity is obvious (see 5. 5). We turn to the sufficiency. Let $X = \mathrm{Hom}(A, Q)$, and let us regard Q itself as a subset of A, rather than merely isomorphic to it (or, what is almost the same thing, select a definite one of the isomorphisms which our hypothesis provides). Let Y be the subset of X containing those y for which $q \varepsilon Q \subset A$ implies $y(q) = q$. Y is evidently compact, and A is homomorphic to a subspace of $C(Y, Q)$. Let $f \varepsilon A$ and suppose $\| f \| = r$. Then there is some $x \varepsilon X$ for which $x(ff^*) = r^2$. By means of a suitable $\alpha \varepsilon \Gamma$ we have $y = \alpha x \varepsilon Y$. But $y(ff^*) = r^2$ as well. Therefore

$$\sup_{y \varepsilon Y} | f(y) | \geqq \| f \|.$$

The reverse inequality is obvivous. Therefore A is algebraically and topologically isomorphic to a subset A' of $C(Y, Q)$ with the k-topology. A' clearly satisfies the conditions of 4. 4, because distinct elements of y have distinct kernel ideals in A. Therefore $A' = C(Y, Q)$, and our theorem is proved.

Suppose we have a concrete BQ^*-algebra A of quaternion-valued functions, and suppose that a basis in Q can be so chosen that for each function f in A, there can be found in A also the four components f_0, f_1, f_2, f_3 of f, where $f = f_0 + f_1 + f_2 + f_3$, to being real-valued, f_1, f_2, f_3 being some real-valued functions times i, j, and k, respectively. It follows that A is closed under the automorphisms $f \rightarrow f^{\mathfrak{i}}$, $f \rightarrow f^{\mathfrak{j}}$, $f \rightarrow f^{\mathfrak{k}}$ where $f^{\mathfrak{i}}(x) = - if(x)i$, etc. These *automorphisms* have the following abstract properties (among others):

5. 81. $\mathfrak{i}\mathfrak{i} = \mathfrak{j}\mathfrak{j} = \textit{identity}, \mathfrak{i}\mathfrak{j} = \mathfrak{j}\mathfrak{i} = \mathfrak{k}$

5. 82. $f + f^{\mathfrak{i}} + f^{\mathfrak{j}} + f^{\mathfrak{k}} = 2(f + f^*)$.

Conversely, we have the following representation theorem.

5. 9. THEOREM. *Let A be a BQ^*-algebra in which there are defined three automorphisms* $\mathfrak{i}$, $\mathfrak{j}$, *and* $\mathfrak{k}$ *satisfying* 5. 81 *and* 5. 82. *Then there exists a compact Hausdorff space Y on which acts a group $G' = \{i', j', k'$, and the identity$\}$ of homeomorphisms satisfying the relations* 5. 81, *leaving a point y_0 of Y fixed. A is isomorphic to the subalgebra B of $C(Y, Q)$ determined by the relations*

5. 91. $f(y)^{\mathfrak{i}} = f(i'y), f(y)^{\mathfrak{j}} = f(j'y), f(y)^{\mathfrak{k}} = f(k'y), f(y_0) = 0$.

(N. B.: The index $\mathfrak{i}$, *etc.* on a quaternion such as q or $f(y)$ denotes the image of that quaternion under the concrete inner automorphism $q \rightarrow - iqi$; but attached to an element f of A it denotes the image under the automorphism abstractly presupposed for A in 5. 9 and satisfying merely the axioms stated. It should be borne in mind that in $f^{\mathfrak{i}}(y)$ the index has been attached to f and not to $f(y)$.)

Moreover (using the same notation for B as for A)

5. 92. $$\| f \| = \sup_{y \varepsilon Y} | f(y) |$$

and

5. 93. $f^*(y) = f(y)^*,\ f^{\mathfrak{i}}(y) = f(y)^{\mathfrak{i}},\ f^{\mathfrak{j}}(y) = f(y)^{\mathfrak{j}},\ f^{\mathfrak{k}}(y) = f(y)^{\mathfrak{k}}.$

Finally, y_0 is isolated if and only if A has a unit.

Proof. We begin by drawing certain consequences from 5. 81 and 5. 82. First, if $f = f^*$ then $f = f^{\mathfrak{i}} = f^{\mathfrak{j}} = f^{\mathfrak{k}}$. This is obvious; when one replaces f^* in 5. 82 by f, one obtains that $4f$ is invariant under all three automorphisms, by 5. 81. Second, each automorphism commutes with the *-operation: to see this, proceed as follows. Replace f in 5. 82 by $g - g^{\mathfrak{i}}$. Then the left side vanishes, yielding $g - g^{\mathfrak{i}} + g^* - g^{\mathfrak{i}*} = 0$. Now apply $\mathfrak{i}$ to $g + g^*$, obtaining $g + g^* = g^{\mathfrak{i}} + g^{*\mathfrak{i}}$. Comparison of these equations gives $g^{\mathfrak{i}*} = g^{*\mathfrak{i}}$.

Define Y as the subset of Hom(A, Q) at which we have $f^{\mathfrak{i}}(y) = f(y)^{\mathfrak{i}}, \cdots,$ $f^{\mathfrak{k}}(y) = f(y)^{\mathfrak{k}}$. G' is the subgroup of the group of homeomorphisms induced by Γ which is defined by 5. 91. Thus 5. 91 and 5. 93 hold, where y_0 is the zero-homomorphism of A. We have now only to prove that 5. 92 holds, which implies incidentally that Y does not consist of y_0 alone (unless A is void).

Let f be given, and suppose $\| f \| = r$. There is an x in Hom(A, Q) such that $| f(x) | = r$. For any g, if $g(x) = 0$ then $gg^*(x) = 0$. Hence $(gg^*)^{\mathfrak{i}}(x) = g^{\mathfrak{i}}(x) g^{*\mathfrak{i}}(x) = g^{\mathfrak{i}}(x) g^{\mathfrak{i}}(x)^* = 0$. Therefore $g^{\mathfrak{i}}(x) = 0$ if $g(x) = 0$; and similar results obtain for $\mathfrak{j}$ and $\mathfrak{k}$. Hence these automorphisms induce inner automorphisms of Q at the various points of Hom (A, Q) (*Cf.* [Jacobson, p. 101, Theorem 15]). If at the point z, all three automorphisms are extensible to Q as the identity, then 5. 82 implies that $g(x) = g(x)^*$ for all g and hence $x \varepsilon Y$. If not, suppose that $g^{\mathfrak{i}}(x) \neq g(x)$ for some g. Then we may suppose that for every g in A, $g^{\mathfrak{i}}(x) = p^* g(x) p$ where p is a *pure* quaternion, for otherwise $\mathfrak{i}^2$ would not be the identity. Let $p = qiq^*$ for some pure q. Let $\gamma \varepsilon \Gamma$ be the inner automorphism $q^* \cdots q$ of Q, and let $z = \gamma x$ in Hom(A, Q). It is easily verified that $g^{\mathfrak{i}}(z) = g(z)^{\mathfrak{i}}$. At the point z consider the inner automorphism of Q induced by $\mathfrak{j}$. As before, suppose $g^{\mathfrak{j}}(z) = u^* g(z) u$ with some pure u. Since $\mathfrak{i}$ does not leave every value at z unchanged, we may conclude that i itself is obtained as one of the values at z. Suppose $h(z) = i$. Inserting h into 5. 82 and evaluating at z, we obtain $u^* i u = -i$. Therefore, a quaternion v may be found such that $v^* i v = i$ and $u = vjv^*$. Let δ be the inner automorphism $v^* \cdots v$, and let $y = \delta z$ in Hom(A, Q). It is easily verified that $g^{\mathfrak{i}}(y) = g(y)^{\mathfrak{i}}$ and $g^{\mathfrak{j}}(y) = g(y)^{\mathfrak{j}}$. Consequently y lies in Y, and 5. 92 is established.

The bi-approximity of A to B is established along the now familiar lines followed in 5. 4. Thus 5. 9 can be established.

This theorem can also be proved by constructing an algebra A_4 of ordered quadruples which (a) contains A and (b) admits the quaternions as scalars. The automorphisms i, j, k are used to define the commutation relation between the elements of A and those of Q. This mode of proof is very tedious because the axioms of BQ^*-algebras for A_4 can only be reduced to those for A by long calculations.

At this point it seems appropriate to consider whether every BQ^*-algebra can not be represented simply as a space $C(Y, Q)$. To put it another way, is the covariance feature 5. 51 of 5. 5 sometimes indispensable, or are the conditions of 5. 8 always fulfilled? The fact is, that not every BQ^*-algebra is representable as a space $C(Y, Q)$.

We shall describe a general method for obtaining BQ^*-algebras which are not representable as spaces $C(Y, Q)$.

Our method is an elaboration of the following principle.

5. 94. LEMMA. *If A is a subalgebra of $C(X, Q)$ and A is isomorphic to $C(Y, Q)$ for some suitable compact Hausdorff space Y, then there exists a continuous mapping of the space X into the orthogonal group Γ.*

Proof. Find f, g, and h in A which correspond to the constant functions i, j, and k in the representation of A as $C(Y, Q)$. Now at each x of X, the triple $f(x)$, $g(x)$, $h(x)$ can, by a unique orthogonal transformation $\alpha(x)$ of the space of pure quaternions, be rotated into the fixed triple of quaternions i, j, k:

5. 941. $$(f(x), g(x), h(x)) = \alpha(x)(i, j, k).$$

This $\alpha(x)$ depends continuously on x and thus provides the desired mapping of X into Γ.

The general method of obtaining examples is now to be set forth.

5. 95. LEMMA. *Let X be a completely regular space containing a closed set z which can be mapped into Γ by a continuous mapping β which cannot be continuously extended to X. Let A be the subalgebra of those f in $C(X, Q)$ defined by the requirement*

5. 951. $$\beta(z_0)(f(z)) = \beta(z)(f(z_0)) \qquad (z \varepsilon Z)$$

where z_0 is some distinguished point of Z arbitrarily selected in advance. Then A admits neither K nor Q as scalars and in particular, is not representable as a space $C(Y, Q)$.

Proof. If A admitted K as scalars, it would be commutative (see Theorem

5. 97 below), but A is surely not commutative because the complement of Z on which the functions in A are subject to no restriction must under the hypothesis of inextensibility of β be a *non-void* open set. This shows incidentally that A isn't void. If A admitted Q as scalars, then as in 5. 94 we could find f, g, h in A which would have the abstract algebraic properties of the quaternions i, j, k. We should then set up a mapping α as we did in 5. 941. For z in Z we would then have

5. 952.
$$\begin{aligned}(z)(i,j,k) &= (f(z), g(z), h(z) \\ &= \beta(z_0)^{-1}\beta(z)(f(z_0), g(z_0), h(z_0))\end{aligned}$$

and consequently

5. 953.
$$\beta(z) = \beta(z_0)\alpha(z)\alpha(z_0)^{-1}.$$

Therefore $\beta(z_0)\alpha(x)\alpha(z_0)^{-1}$ would provide the extension of β from Z to all of X. This provides 5. 94.

5. 96. THEOREM. *There exists a BQ^*-algebra which cannot be represented by a space $C(Y, Q)$, but* modulo *each maximal ideal the residue algebra is Q.*

Proof. Using the terminology of 5. 95, let X be the closed disc ($0 \leqq r \leqq 1$, $0 \leqq \theta < 2\pi$, in polar coordinates), and let Z be the periphery ($r = 1$). Now there exist mappings β of Z into Γ which cannot be extended to all of X. This is because the fundamental group of Γ is not trivial, and because the extension of β to all of X would show that the path given by $\beta(z)$ for $z \varepsilon Z$ was homotopic to the null-path. As an example, we propose $\beta(1, \theta) \varepsilon \Gamma$ to be that rotation of the 2-sphere which corresponds to the inner automorphism $e^{\frac{1}{2}k\theta}(\ \ )e^{-\frac{1}{2}k\theta}$ of the quaternions. This gives a path in Γ which punctures the "plane at infinity" an odd number of times, and at any rate a contemplation of a suitable polytope will convince the reader that this β is not homotopic to 0. Hence condition 5. 951 may be taken as the criterion that f belongs to A, and read

5. 561.
$$f(1, \theta) = e^{\frac{1}{2}k\theta} f(1, 0) e^{-\frac{1}{2}k\theta} \qquad (0 \leqq \theta < 2\pi).$$

Then 5. 95 assures us that A is not representable as a space $C(Y, Q)$.

Each point (r, θ) with $r < 1$ corresponds to one and only one maximal ideal, and all points (r, θ) with $r = 1$ determine one additional ideal. It is easy to see that this accounts for all maximal ideals since a function vanishing at none of these points has an inverse element in A. The values $f(x)$ at any such x, as f ranges over all elements of $\overset{*}{A}$, evidently exhaust Q, whence the residue algebras *modulo* maximal ideals are all isomorphic to Q.

The reason for stressing the residue algebras *modulo* maximal ideals for this example is of course that if for any maximal ideal the residue algebra is not Q, then the possibility that A is representable as a space $C(Y, Q)$ is at once ruled out.

We have been unable to determine whether this example can be represented by 5. 9, but we doubt it.

We end this section with a proposition which shows that it would have been undesirable to limit one's self to BQ^*-algebras admitting complex scalars.

5. 97. THEOREM. *A BQ^*-algebra in which complex scalar multiplication is possible, is commutative.*

Proof. Select any $x \varepsilon X = \mathrm{Hom}(A, Q)$ and an $h \varepsilon A$ such that $x(h) = 1$ and $h = h^*$. Since $x(ih \cdot ih) = x(-h^2) = -1$, $x(ih) = q$ is not a real number in Q. Now $qx(f) = x(ihf) = x(ifh) = x(fih) = x(f)q$ for any f in A. Thus $x(f)$ lies in the commutative subalgebra generated by q. Hence $x(fg - gh) = 0$ for any f and g. Thus A is commutative. The representation of such a ring can therefore be effected by [Arens III, p. 279, Cor. 1].

6. Closed ideals in BQ^*-algebras.

The representation theorems make possible a number of statements about the closed ideals in a BQ^*-algebra.

6. 1. THEOREM. *In a BQ^*-algebra A, every closed ideal is a two-sided ideal, and is an intersection of maximal ideals. The closed ideals of A are in one-to-one correspondence with the family of those compact subsets of $X = \mathrm{Hom}(A, Q)$ which are invariant under the homeomorphisms of X induced by the automorphisms of Q leaving R fixed (see 5. 3). Each closed ideal is closed under the $*$-operation.*

Proof. We shall prove these statements in the reverse of the order stated, using 5. 4.

Let J be a closed ideal. We must show that $f^* \varepsilon J$ if $f \varepsilon J$. However, if $f \varepsilon J$ then $ff^* \varepsilon J$, and so by 5. 09, J contains f^*.

Let $Z \subset X$ be invariant under Γ. Let J_Z be the class of those f in A which vanish on Z. J_Z is evidently a closed two-sided ideal.

Conversely, let J be a closed ideal. For definiteness, suppose J is a right ideal. Let $Z = Z_J$ be the class of those x in X at which every element of J vanishes. Z may be void, but in any case it is closed and invariant under Γ.

We shall show that $J = J_Z$, in the notation of the preceding paragraph, by showing that J is bi-approximate to J_Z, it being clear that J is included in J_Z.

Let x, y in X, and $h \varepsilon J_Z$, be given.

First, suppose x and y lie without Z. Then there exist f and g in J such that $f(x) \neq 0 \neq g(y)$. Now $ff^*(= f^*f)$ and $gg^*(= g^*g)$ belong to J also, and their values at x and y, respectively, are real and positive. By scalar multiplication, we can now obtain f_1 and g_1 such that $f_1(x) = 1$, $g_1(y) = 1$. Now either $y = \alpha x$ for some α, or not. Consider the latter case. Then we can find, as in the proof of 5. 4, h_x and h_y in A such that $h_x(x) = h_y(y) = 1$ and $h_x(y) = h_y(x) = 0$. The element $f_1h_xh + g_1h_yh$ belongs to J. It has the same values at x and y, respectively, as h has. In the former case, $h' = f_1h$. This element lies in J; and

$$h'(x) = f_1(x)h(x) = h(x),$$

$$h'(y) = h'(\alpha x) = \alpha(h'(x)) = \alpha(h(x)) = h(\alpha x) = h(y).$$

Second, suppose $y \varepsilon Z$. We can dismiss the case that also $x \varepsilon Z$. We obtain f_1 as before, and form $h' = f_1h$ in J. Then $h'(x) = h(x)$, and $h'(y) = 0 = h(y)$.

We conclude that $J = J_Z$, by 4. 3.

The correspondence referred to in 6. 1 is that between J_Z and Z_J. It will be observed that it is inclusion-reversing, as is to be expected in these matters. Therefore, the maximal ideals correspond to the smallest invariant sets, i. e., the orbits of the individual points of X under Γ. Since each Z_J is a union of some of the latter, each J_Z is an intersection of the former.

It is already abundantly clear that each right ideal is also a left ideal, but this can also be demonstrated as follows. Let J be a right ideal. Let $f \varepsilon J$ and $g \varepsilon A$. Then $(gf)^* = f^*g^*$ lies in J because f^* does. Hence $gf \varepsilon J$. This completes the proof of 6. 1.

6. 2. COROLLARY. *If X is a compact Hausdorff space, then the closed ideals of $C(X, Q)$, with the k-topology, are in one-to-one inclusion-reversing correspondence with the compact subsets of X.*

This theorem may be extended to completely regular spaces X. The sets that correspond to closed ideals are, however, still compact. We leave the proofs to the reader. And then the results may, of course, be specialized to the case of K or R, rather than Q.

In finite-dimensional BQ^*-algebras, all ideals are linear subspaces and hence are closed. Therefore, in the finite-dimensional case, all ideals are two-sided. On the other hand, one of the simplest infinite-dimensional BQ^*-

algebras does have ideals which are not two-sided and hence also, *a fortiori*, ideals which are not closed.

6.3. EXAMPLE. *If X is the unit interval, then the BQ^*-algebra $C(X, Q)$ contains a right ideal which is not two-sided* (*Cf.* [Kaplansky, p. 180]).

The ideal we wish to exhibit is the principal right ideal J generated by the function

$$g(x) = x(\cos 1/x + i \sin 1/x) \qquad 0 < x \leqq 1.$$

$$g(0) = 0.$$

To show that J is not two-sided, we need only attempt to find an f in $C(X, Q)$ such that $jg = gf$. If this were possible, a simple calculation shows that, for $0 < x \leqq 1$ we must have

$$f(x) = j \cos 2/x - k \sin 2/x.$$

Hence no definition of $f(0)$ can make f belong to $C(X, Q)$. Therefore jg does not belong to J, and J is merely a right ideal.

7. A remark on Banach *-algebras.

This section deals with a detail on Banach *-algebras, i.e., Banach algebras A with complex scalars, in which there is defined a semilinear operation (*) satisfying 5.01 and also

7.01. $$(cf) = c^*f^* \qquad (c \text{ complex}).$$

This section is otherwise unrelated to the preceding theory. Our purpose is to prove the following theorem.

7.1. THEOREM. *Let A be a Banach *-algebra with a unit, in which 5.04 holds and in which, for all f,*

7.2. $$\| ff^* \| = \| f \|^2.$$

There can be no element f in A such that $ff^* + 1 = 0$.

Before proceeding to the proof of 7.1, we indicate the origin of this question. In [Gelfand and Neumark, p. 196 (footnote)] it is conjectured that from the hypothesis of 7.1 above it follows that $ff^* + 1$ have a unique two-sided inverse for every f (in the commutative case, this is known to be true). From this conjecture, 7.1 would follow. Hence the conjecture is at least slightly substantiated by 7.1.

Proof of 7.1. We shall show first, that if $ff^* = -1$ then the smallest circle in the complex plane with center at the origin and surrounding the spectrum of f contains the unit circle. For the radius of this circle is given by

$$r = \overline{\lim} \| f^n \|^{1/n} \tag{7.3}$$

(*Cf.* [Arens III, first part of proof of Lemma 2] or, of course, [Gelfand]). Now the radius of a similar circle for f^* is clearly the same as r, for algebraic reasons. By 7.2,

$$\| f^n \|^2 = \| f^n (f^*)^n \| = \| 1 \| = 1,$$

hence

$$\| f^n \|^{1/n} = 1.$$

By letting n approach infinity and observing 7.3, we obtain $r \geqq 1$. As a matter of fact, $r \leqq \| f \|$, $\| f^* \| = 1$ because the spectrum of any element is confined to a disc about the origin whose radius is the norm. Consequently $r = 1$.

We proceed to another calculation. Let c be any complex number of absolute value 1. Then $(f - c)(f^* - c^*) = ff^* - cf^* - c^*f + 1$. Since $ff^* + 1 = 0$ we obtain

$$\| f - c \|^2 = \| (f - c)(f^* - c^*) \| \leqq \| f \| + \| f^* \| = 2.$$

Therefore the spectrum of f is within a distance of $2^{1/2}$ of *every* point on the unit circle. This contradicts the minimum property already established for the unit circle. Thus 7.1 is proved.

UNIVERSITY OF CALIFORNIA AT LOS ANGELES.

BIBLIOGRAPHY.

L. Alaoglu, "Weak topologies of normed linear spaces," *Annals of Mathematics*, vol. 41 (1940), pp. 252-267.

R. Arens, (I), "Topologies for spaces of transformations," *Annals of Mathematics*, vol. 47 (1946), pp. 480-495.

———, (II), "Topological division algebras," *Bulletin of the American Mathematical Society*, vol. 53 (1947), pp. 623-630.

———, (III), "Representation of *-algebras," *Duke Mathematical Journal*, vol. 14 (1947), pp. 269-282.

R. Arens and I. Kaplansky, "Topological representation of algebras," *Transactions of the American Mathematical Society*, vol. 63 (1948), pp. 457-481.

G. Birkhoff, "Lattice ordered groups," *Annals of Mathematics*, vol. 43 (1942), pp. 298-331.

J. A. Clarkson, "A characterization of C-spaces," *Annals of Mathematics*, vol. 48 (1947), pp. 845-50.

L. E. Dickson, *Linear Algebras*, Cambridge Mathematical Tracts 16, Cambridge (1914).

N. Dunford and I. E. Segal, "Semigroups of operators and the Weierstrass theorem," *Bulletin of the American Mathematical Society*, vol. 52 (1946), pp. 911-914.

I. Gelfand, "Normierte Ringe," *Recueil Mathématique*, N. S., vol. 9 (51), (1941), pp. 3-24.

I. Gelfand and M. Neumark, "On the imbedding of normed rings into the ring of operators in Hilbert space," *Recueil Mathématique*, N. S., vol. 12 (54), (1943), pp. 197-213.

N. Jacobson, *Theory of rings*, Mathematical Surveys 2, American Mathematical Society, New York (1943).

S. Kakutani, "Concrete representation of abstract (M)-spaces," *Annals of Mathematics*, vol. 42 (1941), pp. 994-1024.

I. Kaplansky, "Topological rings," *American Journal of Mathematics*, vol. 69 (1947), pp. 153-183.

H. Lebesgue, "Sur l'approximation des fonctions," *Bulletin des Sciences Mathématiques*, vol. 22 (1898), pp. 278-287.

M. H. Stone, "Applications of the theory of Boolean rings to general topology," *Transactions of the American Mathematical Society*, vol. 41 (1937), pp. 375-481.

MULTIPLICATIVE DIOPHANTINE EQUATIONS IN QUATERNIONS.*

By E. ROSENTHALL.

---

1. This note is concerned with the resolution of multiplicative equations in integral quaternions. We consider only Lipschitz integral quaternions. $T = t_1 + it_2 + jt_3 + kt_4$ with rational integral coordinates t_i, where i, j, k are the familiar Hamiltonian units. Conjugate and norm of T are defined as usual: $\bar{T} = 2t - T$, $N(T) = \bar{T}T = \sum t_i^2$; $\mathfrak{R}[T]$ will represent the rational part t_1 of T. Except for the quaternion units i, j, k, all small Latin letters with or without subscripts will denote rational integers; large capital letters $A, B, \cdots$ will represent integral quaternions; their corresponding coordinates are designated by a_1, a_2, a_3, a_4, and so on. We say that A is even or odd according as $N(A)$ is even or odd, and hence according as $a_1 + a_2 + a_3 + a_4$ is even or odd. If $(a_1, a_2, a_3, a_4) = 1$, A is said to be primitive. The Greek letter π will represent any one of the five even quaternions $1 + i$, $1 + j$, $1 + k$, $(1 + i)(1 + j)$, $(1 + i)(1 + k)$.

The main object of this paper is to obtain the complete integral parametric representation in integral formulas of the equation $XY = ZW$. The resolution of this equation will be based on Theorems 1 and 2. Theorem 1 is interesting in itself and generalizes the problem of finding all integral quaternions which are doubly divisible by a given integral quaternion [4]. Theorem 2 is not new [2, p. 7], but the proof given here shows that the result is an immediate consequence of the factorization of primitive quaternions.

2. THEOREM 1. *The complete integral solution of the system*

$$TW = ZU, \qquad N(T) = N(U), \qquad T \text{ odd and primitive,} \tag{1}$$

is given by

$$W = \bar{T}X + \bar{Y}U, \qquad Z = X\bar{U} + T\bar{Y}, \tag{2}$$

where X and Y are arbitrary.

Substituting in (1) the value of W from (2) we get $ZU = T\bar{T}X + T\bar{Y}U = (X\bar{U} + T\bar{Y})U$, $Z = X\bar{U} + T\bar{Y}$. Hence it suffices to show that the linear system, obtained by equating coordinates in (2),

---

* Received June 12, 1948.

$$
(3)\qquad
\begin{aligned}
t_1x_1 + t_2x_2 + t_3x_3 + t_4x_4 + u_1y_1 + u_2y_2 + u_3y_3 + u_4y_4 &= w_1,\\
-t_2x_1 + t_1x_2 + t_4x_3 - t_3x_4 + u_2y_1 - u_1y_2 - u_4y_3 + u_3y_4 &= w_2,\\
-t_3x_1 - t_4x_2 + t_1x_3 + t_2x_4 + u_3y_1 + u_4y_2 - u_1y_3 - u_2y_4 &= w_3,\\
-t_4x_1 + t_3x_2 - t_2x_3 + t_1x_4 + u_4y_1 - u_3y_2 + u_2y_3 - u_1y_4 &= w_4,
\end{aligned}
$$

is always solvable for integers x_i, y_i when T, W, U satisfy (1). System (3) will be solvable in integers if the g. c. d. of all the four-rowed determinants in the matrix of the coefficients is equal to the g. c. d. k of all the four-rowed determinants in the augmented matrix (M) [6, p. 10]. Let (a, b, c, d) denote the determinant formed from the columns in position a, b, c, d in the augmented matrix, and let $m = N(T)$. Then $(1, 2, 3, 4) = m^2$, and so k can only contain factors of m. Also $(1, 3, 6, 8) = m[m - (u_1^2 + u_3^2 + t_1^2 + t_3^2)] = ma$ and $(1, 3, 5, 7) = m[m - (t_1^2 + t_3^2 + u_2^2 + u_4^2)] = mb$. Let $p \mid m$, where p is a rational prime, and let $p \mid a$ and $p \mid b$; then $p \mid a + b$, $p \mid t_1^2 + t_3^2$. Similarly if $p \mid c$ and $p \mid d$ in $(1, 2, 5, 6) = m[m - (t_1^2 + t_2^2 + u_3^2 + u_4^2)] = mc$, and in $(3, 4, 5, 6) = m[m - (t_1^2 + t_2^2 + u_1^2 + u_2^2)] = md$, then $p \mid t_1^2 + t_2^2$; and if $p \mid e$ and $p \mid f$ in $(2, 3, 5, 8) = m[m - (u_1^2 + u_4^2 + t_1^2 + t_4^2)] = me$ and in $(1, 4, 5, 8) = m[m - (u_2^2 + u_3^2 + t_1^2 + t_4^2)] = mf$, then $p \mid t_1^2 + t_4^2$. Hence $p \mid 2t_1^2 + N(T)$, $p \mid t_i$ $(i = 1, 2, 3, 4)$, $p = \pm 1$, since T is primitive and therefore $k \mid m$. We have only to show now that all the four-rowed determinants are divisible by m.

Note that the elements of the successive columns in (M), taken in order of the rows, are respectively the components of the quaternions $\bar{T}$, $\bar{T}i$, $\bar{T}j$, $\bar{T}k$, U, $-iU$, $-jU$, $-kU$, W. We observe that the determinant

$$
(4)\qquad d = \begin{vmatrix} t_1 & t_2 & a_1 & b_1 \\ -t_2 & t_1 & a_2 & b_2 \\ -t_3 & -t_4 & a_3 & b_3 \\ -t_4 & t_3 & a_4 & b_4 \end{vmatrix}
$$

is equal to $m(b_1a_2 - a_1b_2) + (t_1^2 + t_2^2)(B\bar{A})_i + (t_2t_3 - t_1t_4)(B\bar{A})_j + (t_2t_4 + t_1t_3)(B\bar{A})_k$, where $(B\bar{A})_i$ is the coefficient of i in the product $B\bar{A}$. Since $\bar{T}iT = -im + 2i(t_1^2 + t_2^2) + 2j(t_2t_3 - t_1t_4) + 2k(t_1t_3 + t_2t_4)$ it follows that

$$
(5)\qquad 2d \equiv -\mathfrak{R}[\bar{T}iT\,B\bar{A}] \pmod{m} \text{ or } 2d \equiv \mathfrak{R}[\bar{T}iT\,A\bar{B}] \pmod{m}.
$$

If $B = W$ and $A = \theta U$ or $\bar{T}\theta$, where θ is a unit then $2d \equiv -\mathfrak{R}[\bar{T}iTW\bar{U}\theta] = \mathfrak{R}[\bar{T}iZU\bar{U}\theta] \equiv 0 \pmod{m}$ or $2d \equiv \mathfrak{R}[\bar{T}iT\bar{T}\theta\bar{W}] \equiv 0 \pmod{m}$. Hence all the four-rowed determinants with columns 1, 2 and 9 are divisible by m.

Again if $B = \theta_1 U$, $A = \theta_2 U$, then $2d \equiv -\mathfrak{R}[\bar{T}iT\theta_1 U\bar{U}\theta_2] \equiv 0 \bmod m)$ and if $B = \theta_1 U$ and $A = \bar{T}\theta_2$ then $2d \equiv \mathfrak{R}[\bar{T}iT\bar{T}\theta_2\bar{U}\bar{\theta}_1] \equiv 0 \bmod m)$; hence all four-rowed determinants of (M) with columns 1 and 2 are divisible by m.

If in (4) we replace $\bar{T}$ by $\bar{T}j$, (5) remains unchanged except for sign; but this substitution replaces columns 1 and 2 of (M) by column 3 and the negative of column 4 of (M). Hence all four-rowed determinants of (M) with columns 3 and 4 are divisible by m.

In (4) replace T by U and change the signs of the elements in the last three rows. This gives a four-rowed determinant whose first two columns are columns 5, 6 of (M), and (5) becomes

$$(6) \qquad 2d \equiv \mathfrak{R}[\bar{U}iU \cdot B\bar{A}] (\bmod m) \text{ or } 2d \equiv -\mathfrak{R}[\bar{U}iU \cdot AB] (\bmod m).$$

Put $\bar{B} = W$ and $\bar{A} = \theta U$ or $\bar{T}\theta$; then (6) gives $2d \equiv -\mathfrak{R}[\bar{U}iU\bar{U}\bar{\theta}W] \equiv 0$ $(\bmod m)$ or $2d \equiv \mathfrak{R}[\bar{U}iU\bar{W}\bar{T}\theta] = \mathfrak{R}[\bar{U}iU\bar{U}\bar{Z}\theta] \equiv 0$ $(\bmod m)$. If $\bar{A} = \bar{T}\theta_1$, $\bar{B} = \bar{T}\theta_2$, then $2d \equiv \mathfrak{R}[\bar{U}iU\bar{\theta}_2 T\bar{T}\theta_1] \equiv 0 \bmod m)$, and if $\bar{A} = \bar{T}\theta_1$ and $\bar{B} = \theta_2 U$, then $2d \equiv \mathfrak{R}[\bar{U}iU\bar{U}\bar{\theta}_2\bar{T}\theta_1] \equiv 0$ $(\bmod m)$. Hence all four-rowed determinants of (M) containing columns 5 and 6 are divisible by m; similarly for the determinants with columns 7 and 8, as is seen by replacing U by $-jU$.

If in (4) we replace the first two columns by columns 1 and 3 of (M), then d satisfies $2d \equiv -\mathfrak{R}[\bar{T}jT \cdot B\bar{A}]$ $(\bmod m)$ or $2d \equiv \mathfrak{R}[\bar{T}jT \cdot A\bar{B}]$ $(\bmod m)$; and if in (4) we replace the first two columns by columns 1 and 4 of (M), then d satisfies

$$2d \equiv -\mathfrak{R}[\bar{T}kT \cdot B\bar{A}] \ (\bmod m) \text{ or } 2d \equiv \mathfrak{R}[\bar{T}kT \cdot A\bar{B}] \ (\bmod m).$$

We now proceed with each of these two sets of formulas for $2d$ as we did with (5). If in these formulas we replace successively T by $-iT$, U, $-iU$, we obtain (except possibly for sign) from the first set of formulas the values of $2d (\bmod m)$ when the first two columns of (4) are respectively replaced by columns 2 and 4 of (M), by columns 5 and 7 of (M), by columns 6 and 8 of (M), and from the second set we get the values of $2d (\bmod m)$ when these two columns of (4) are respectively replaced by columns 2 and 3 of (M), by columns 5, 8 of (M), by columns 6, 7 of (M). Now in each of these formulas make the same substitutions for A and B as we made in (5) and (6). In all cases we get $d \equiv 0$ $(\bmod m)$, and this completes the proof that all four rowed determinants are divisible by m, and hence the theorem.

A result due to Olson[4] giving all integral quaternions doubly divisible by a given integral quaternion T can be deduced from Theorem 1.

COROLLARY 1. *All integer solutions of* $TW = ZT$, *T odd and primitive are given by*

$$Z = N\bar{T} + 2a, \qquad W = \bar{T}N + 2a$$

where a and N are arbitrary.

For Z, W satisfy $W = \bar{T}X + \bar{Y}T$, $Z = X\bar{T} + T\bar{Y}$. Replace X by $Y + N$ and replace the equal rational integers $T\bar{Y} + Y\bar{T}$, $\bar{T}Y + \bar{Y}T$ by $2a$.

In the case T even and primitive we can write [3, p. 304] $T = S\pi$ where S is odd; also there is an integer V such that $\pi W = V\pi$, where W is obtained from V by known sign changes and interchanges of coordinates of V, depending on the value of π selected. Since $TW = ZT$, it follows that $SV = ZS$, and so $Z = N\bar{S} + 2a$, $V = \bar{S}N + 2a$.

3. THEOREM 2. *All odd integral solutions of the equation*

$$N(X) = uv \tag{7}$$

are given by

$$X = mAB, \qquad u = mA\bar{A}, \qquad v = mB\bar{B}, \tag{8}$$

where m, A, B are odd.

Let $X = pY$, Y primitive; then $p^2Y\bar{Y} = uv$ and we can write $p^2 = ab$, $Y\bar{Y} = cd$, $u = ad$, $v = cb$. Since Y is primitive, c odd and $c \mid N(Y)$, then [5, Theorem 1, p. 488] Y has a right divisor of norm c, $Y = DC$, $c = C\bar{C}$ and hence $D\bar{D} = d$. Also from $p^2 = ab$ we can write $p = mnq$, $a = mn^2$, $b = mq^2$. Hence

$$X = mnqDC, \qquad u = mn^2D\bar{D}, \qquad v = mq^2C\bar{C},$$

and writing B, A respectively for the products qC, nD gives the required result.

If m, A, B are arbitrary then (8) gives all the solutions of (7); for put $Y = \pi Z$, Z odd, then $Z\bar{Z} = u_1v$, where $u = 2u_1$ when $N(\pi) = 2$, and $Z\bar{Z} = u_1v_1$ or u_2v, where $u = 2u_1$, $v = 2v_1$, $u = 4u_2$ when π is $(1 + i)(1 + j)$ or $(1 + i)(1 + k)$. Since Z is odd, from $Z\bar{Z} = u_1v$ or u_2v by (8) we get $Y = \pi mEB$, $u = 2mE\bar{E}$ or $4mE\bar{E}$ respectively, $v = mB\bar{B}$. Replacing πE by A gives the required form (8). Finally from $Z\bar{Z} = u_1v_1$ we get $Y = \pi mEF$, $u = 2mE\bar{E}$, $v = 2mF\bar{F}$, which takes the form (8) if we replace πE by $A(1 + j)$ and $(1 + j)F$ by B when $\pi = (1 + i)(1 + j)$, and πE by $A(1 + k)$ and $(1 + k)F$ by B when $\pi = (1 + i)(1 + k)$.

4. We can now reduce the resolution of $XY = ZW$ to that of system (1) by using the conditions for quaternions to have common divisors of odd norm as given by G. Pall [5].

THEOREM 3. *All odd integer solutions of*

$$XY = ZW \tag{9}$$

are given by

$$\begin{aligned} X &= cS\bar{S}FPABC, \qquad Z = eC\bar{C}F(U\bar{N} + P\bar{V})RST, \\ Y &= eT\bar{T}\bar{C}\bar{B}\bar{A}(\bar{P}U + \bar{V}N), \qquad W = cB\bar{B}\bar{T}\bar{S}\bar{R}NE, \end{aligned} \tag{10}$$

where $A\bar{A} = R\bar{R}$, $P\bar{P} = N\bar{N}$ and all parameters are odd except that U and V must be of opposite parity.

Put

$$X = xX_1, \qquad Y = yY_1, \qquad Z = zZ_1, \qquad W = wW_1, \tag{11}$$

where X_1, Y_1, Z_1, W_1 are each primitive. Let k and l, both necessarily odd, denote the g. c. d. of the coordinates of the product X_1Y_1 and the product Z_1W_1 respectively. Then [5, Th. 5, p. 491] we can write $X_1 = X_2K$, $Y_1 = \bar{K}Y_2$, $N(K) = k$ where the product X_2Y_2 is primitive; similarly $Z_1 = Z_2L$, $W_1 = \bar{L}W_2$, $N(L) = l$, where the product Z_2W_2 is primitive. Thus (8) becomes $xyK\bar{K}X_2Y_2 = zwL\bar{L}Z_2W_2$. Hence

$$xyK\bar{K} = zwL\bar{L} \text{ and } X_2Y_2 = Z_2W_2. \tag{12}$$

The solutions of the first equation in (12) satisfy

$$x = abc, \quad y = def, \quad z = aen, \quad w = dhc; \quad K\bar{K} = ghn, \quad L\bar{L} = gbf,$$

and by a repeated application of Theorem 2, since K is primitive, all the solutions of $K\bar{K} = ghn$ are given by $K = ABC$, $g = A\bar{A}$, $h = B\bar{B}$; $n = C\bar{C}$; similarly all the solutions of $L\bar{L} = gbf$ are given by $L = RST$, $g = R\bar{R}$, $b = S\bar{S}$, $f = T\bar{T}$. Equating the two values of g shows that $A\bar{A} = R\bar{R}$. Hence all the odd integers satisfying (12) with K and L primitive are contained in

$$x = caS\bar{S}, \quad y = deT\bar{T}, \quad z = aeC\bar{C}, \quad w = dcB\bar{B}; \quad K = ABC, \quad L = RST, \tag{13}$$

where $A\bar{A} = R\bar{R}$ and all parameters are odd.

Now consider the second equation in (12). Let m, necessarily odd, divide both $N(Y_2)$ and $N(W_2)$. Then since X_2Y_2 is primitive it follows [5, p. 489, Corollary 1′] that any right divisor of norm m of X_2Y_2 is also a right divisor of norm m of Y_2, and conversely; that is, Z_2W_2 and Y_2 have the same right divisors of norm m. Since Z_2W_2 is primitive, Z_2W_2 and W_2 also have the same right divisors of norm m; hence any right divisor of norm m of Y_2 is also a right divisor of norm m of W_2, and conversely.

Therefore, if $(N(W_2), N(Y_2)) = m$, we necessarily have $Y_2 = ME$, $W_2 = NE$, $(N(M), N(N)) = 1$, $N(E) = m$, and (12_2) reduces to $X_2M = Z_2N$. Similarly, if $(N(X_2), N(Z_2)) = p$, it is necessary that $X_2 = FP$, $Z_2 = FQ$, where $(N(P), N(Q)) = 1$, and therefore $PM = QN$, $(N(P), N(Q)) = 1$, $(N(M), N(N)) = 1$, hence $N(P) = N(N)$. Thus the resolution of (12_2) is reduced to that of

$$PM = QN, \qquad N(P) = N(N). \tag{14}$$

Since P is necessarily odd and primitive, it follows from (14) by Theorem 1 that $M = \bar{P}U + \bar{V}N$, $Q = U\bar{N} + P\bar{V}$. Hence all the odd solutions of $X_2Y_2 = Z_2W_2$, both sides primitive, are contained in the set

$$X_2 = FP, \quad Y_2 = (\bar{P}U + \bar{V}N)E, \quad Z_2 = F(U\bar{N} + P\bar{V}), \quad W_2 = NE, \tag{15}$$

where $P\bar{P} = N\bar{N}$, and all parameters are odd except that U and V are of opposite parity. Substituting (15) and (13) into (11) and absorbing a into F and d into E yields (10).

5. The solution (10) of (9), though complete, is not stated explicitly, since the parameters A, R, P, N must satisfy the systems $A\bar{A} = R\bar{R}$, R odd and $P\bar{P} = N\bar{N}$, P odd. Each of these systems, which are of the form $X\bar{X} = Y\bar{Y}$, X odd, can be solved explicitly in terms of quaternions by an application of Theorem 1. In fact we establish

THEOREM 4. *All solutions of $X\bar{X} = Y\bar{Y}$ are given by*

$$fX = MU + UN, \qquad fY = M\bar{U} - UN, \tag{16}$$

where N is pure and it suffices to take $U = V\pi$, where V is odd and primitive and f is 2 or 4.

Put $P = X - Y$, $Q = X + Y$ and equation becomes $P\bar{Q} = -Q\bar{P}$. Since Q is even we can write $Q = eV\pi$, where V is odd and primitive. If also $N(\pi) = 2$ we can put $P = W\pi$. Hence in this case $W\bar{V} = -V\bar{W}$, and so, by Theorem 1, $W = KV + VL$ and $\bar{W} = -(\bar{V}K + L\bar{V})$. It follows that $KV + VL = -(\bar{K}V + V\bar{L})$, $k_1 = -l_1$ and it therefore suffices to take K and L pure. Hence $X = \frac{1}{2}(MV\pi + VL\pi)$, $Y = \frac{1}{2}(\bar{M}V\pi - VL\pi)$, where we have replaced $K + e$ by M or, since there exists a pure quaternion N satisfying

$L\pi = \pi N$, replacing $V\pi$ by U we can write the solutions as $2X = MU + UN$, $2Y = \bar{M}U - UN$, where N is pure.

In the case $Q = eV\pi$, where $N(\pi) = 4$, P will have a divisor of norm 2, and proceeding as above we get $4X = MU + UN$, $4Y = M\bar{U} - UN$, where N is pure and we have replaced $K + 2e$, $V\pi$ by M, U respectively.

The solutions (16) are equivalent to those given by E. T. Bell [1, p. 95] in terms of rational integral parameters. For, if we put

$$U = a_1 + ia_2 + ja_3 + ka_4,$$

$$2M = 2a - i(b_{34} + b_{12}) + j(b_{13} - b_{24}) - k(b_{23} + b_{14}),$$

$$2N = i(b_{34} - b_{12}) - j(b_{24} + b_{13}) + k(b_{23} - b_{14}),$$

the solutions (16) take the form

$$fx_i = -\sum_{j=1}^{i-1} a_j b_{j\,i} + aa_i + \sum_{j=1}^{4-i} a_{i+j} b_{i\,i+j}, \tag{17}$$

$$fy_i = -\sum_{j=1}^{i-1} a_j b_{j\,i} + aa_i - \sum_{j=1}^{4-i} a_{i+j} b_{i\,i+j}, \tag{18}$$

where it suffices to take $b_{12} \equiv b_{34} \pmod 2$, $b_{13} \equiv b_{24} \pmod 2$, $b_{14} \equiv b_{23} \pmod 2$. We now show that it suffices to take $f = 2$, but we must then give up the congruence conditions on b_{ij}. When $f = 4$ we have a even and also $U = V\pi$, where $N(\pi) = 4$ and V odd and primitive. Hence all the a_i are odd and it follows that b_{12}, b_{13}, b_{14} are all even or two of them are odd. In the former case replace a by $2a$, b_{ij} by $2b_{ij}$ in (17) and (18) and divide by 2; the result is (17) and (18) with $f = 2$. In the second case let, for example, b_{12}, b_{13}, hence b_{34}, b_{24}, be odd and let b_{14}, hence b_{23}, be even; then make the replacements $a = 2a$, $b_{12} = 2b_{12} + a_3a_4$, $b_{13} = 2b_{13} - a_2a_4$, $b_{24} = 2b_{24} + a_1a_3$, $b_{34} = 2b_{34} - a_1a_2$, $b_{14} = 2b_{14}$, $b_{23} = 2b_{23}$. If we divide by 2, we obtain (17), (18) with $f = 2$.

For the complete solution of $X\bar{X} = Y\bar{Y}$, X odd, it remains to specify the range of the parameters a_i, b_{ij} having the property that the expressions on the right hand side of (17) (hence, of (18)) become even and their sum $\equiv 2 \pmod 4$. The necessary and sufficient congruence relations for these conditions to hold can be stated as follows:

Let p, q, r, s be some permutation of 1, 2, 3, 4, and let $b_{ij} = -b_{ji}$.

(i) The a_i are not all odd;

(ii) if a_p even, and a_q, a_r, a_s odd then $b_{pq} + b_{pr} + b_{ps}$ is even,

$b_{qr} \equiv b_{rs} \equiv b_{sq} \pmod 2$ and if $a_q \equiv a_r \equiv \pm a_s \pmod 4$ then the remaining parameter a is given by $a_q \equiv 2 + b_{pq} + 2b_{pr} \pm b_{ps} \pmod 4$;

(iii) if a_p, a_q even, a_r, a_s odd, $a_r \equiv \pm a_s \pmod 4$, then $b_{pr} + b_{ps}$, $b_{qr} + b_{ps}$, $a + b_{rs}$ are even and if $a_p + a_q \equiv 0 \pmod 4$ then $2a \equiv 2 + b_{pr} + b_{qr} \pm b_{ps} \pm b_{qs} \pmod 4$, but if $a_p + a_q \equiv 2 \pmod 4$ then $2b_{pq} \equiv 2 + b_{pr}$ $b_{qr} \pm b_{ps} \pm b_{qs} \pmod 4$;

(iv) if a_p odd and a_q, a_r, a_s even then b_{pq}, b_{pr}, b_{ps} are even and if $a_q \equiv a_r \equiv a_s \pmod 4$ then $a \equiv 2 + b_{pq} + b_{pr} + b_{ps} \pmod 4$, but if $a_q \equiv a_r \not\equiv a_s \pmod 4$ then $a \equiv b_{pq} + b_{pr} + b_{ps} + 2(b_{qs} + b_{rs}) \pmod 4$.

To obtain (i), note that if $a_i \equiv \pm 1 \pmod 4$, then

$$2(\sum x_i) \equiv \pm(a - b_{12} - b_{13} - b_{14}) \pm (b_{12} + a - b_{23} - b_{24}) \pm (b_{13} + b_{23} + a - b_{34}) \pm (b_{14} + b_{24} + b_{34} + d) \pmod 4,$$

and also that the expressions in each bracket are even, being respectively congruent to $2x_1$, $2x_2$, $2x_3$, $2x_4 \pmod 2$; hence it suffices to take only the positive signs. Therefore $\sum x_i \equiv 2a \pmod 2$, and so all the a_i cannot be odd.

For (ii), (iii) and (iv) it suffices to obtain the condition for a definite selection of p, q, r, s, since (17) and (18) remain unaltered when any two subscripts are interchanged. It suffices to take all the a_i not even, since it suffices to have $U = V\pi$, V odd and primitive.

6. As an application of Theorem 2 we can obtain the complete integral solution of the multiplicative equation $xX\bar{X} = yY\bar{Y}$ in integral formulas. Put $X = wW$, $Y = zZ$ where W and Z are primitive; then this equation becomes $xw^2(W\bar{W}) = yz^2(Z\bar{Z})$ and hence we can write $x = ars^2c$, $w = glmn$, $y = aeln^2$, $z = grst$, $W\bar{W} = dert^2$, $Z\bar{Z} = dlm^2c$. From the last two equations, by Theorem 2, $W = AGFE$, $e = E\bar{E}$, $r = F\bar{F}$, $t^2 = G\bar{G}$, and $Z = CKHJ$, $l = H\bar{H}$, $c = J\bar{J}$, $m^2 = K\bar{K}$ where $d = A\bar{A} = C\bar{C}$. Hence the complete solution can be written as

$$(19) \quad X = AGL\bar{H}HF, \quad Y = CKHF\bar{F}M, \quad x = a\bar{F}M\bar{M}F, \quad y = aL\bar{H}H\bar{L},$$

where we have replaced the products nE, sJ, gA, gC, mG, tK by L, M, A, C, G, K respectively. If now we put $L\bar{H} = P$, $HF = Q$, $\bar{F}M = R$, $AG = S$, $CK = T$ then all the numbers (19) are included in

$$(20) \quad X = SPQ, \quad Y = TQR, \quad x = aR\bar{R}, \quad y = aP\bar{P}, \text{ where } S\bar{S} = T\bar{T}.$$

By direct substitution we see that all the numbers (20) are solutions of the given equation. Hence (20) is the complete solution of $xX\bar{X} = yY\bar{Y}$.

We conclude with a discussion of the equation $X^2 = ZW$. The complete solution of this equation would be given by (10), where the parameters are subject to the condition obtained by equating the expressions for X and Y. This leads to an equation which cannot be handled with facility, but some progress can be made if in addition we require that X^2 be primitive. For if $X^2 = ZW$, we can write $N(X) = klm$, $N(Z) = kl^2$, $N(W) = km^2$, where $(l, m) = 1$. Since X^2 is primitive, the left divisors of norm kl of X and Z are the same, and the right divisors of norm km of X and W are the same; hence we can put $X = AP$, $Z = AQ$, $X = RB$, $W = UB$, where $N(A) = kl$, $N(B) = km$, $N(P) = N(U) = m$, $N(R) = N(Q) = l$. Therefore $PR = QU$, and by Theorem 1, $R = \bar{P}S + TU$, $Q = S\bar{U} + PT$ where $P\bar{P} = U\bar{U}$. Also $AP = RB$, and R must be a left divisor of norm l of A; hence $A = RV$, $B = VP$. Hence all the solutions of $X^2 = ZW$, X^2 odd and primitive, are contained in the set

$$X = (\bar{P}S + TU)VP, \quad Z = (\bar{P}S + TU)V(S\bar{U} + PT), \quad W = UVP, \tag{21}$$

where $P\bar{P} = U\bar{U}$, and all parameters are odd except that S and T are of opposite parity.

The set (21), however, is not the complete primitive solution since the precise range of the parameters necessary and sufficient to isolate these solutions is not available.

THE INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

1. E. T. Bell, "Separable diophantine equations," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 86-102.

2. L. E. Dickson, "On the theory of numbers and generalized quaternions," *American Journal of Mathematics*, vol. 46 (1924), pp. 1-16.

3. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford (1938).

4. H. L. Olson, "Doubly divisible quaternions," *Annals of Mathematics*, vol. 31 (1930), pp. 371-374.

5. Gordon Pall, "On the arithmetic of quaternions," *Transactions of the American Mathematical Society*, vol. 47 (1940), pp. 487-500.

6. Th. Skolem, "'Diophantische Gleichungen," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 5 (1938), no. 4.

DIVISOR ALGEBRAS.* [1]

By Israel Nathan Herstein.

Introduction. The object of this paper is to characterize the divisor algebras of algebraic function fields of degree of transcendence one over algebraically-closed constant fields by a "natural" system of axioms.

To any algebraic function field of degree of transcendence one over a constant field there has been associated a set of objects called the divisors of the function field. [1, chapters 2 & 3].[2] For these divisors multiplication, equivalence and dependence have been introduced so that under this multiplication the divisors form a free Abelian group, and equivalence classes of divisors form projective geometries under the dependence.

We will consider a free Abelian group—the group of divisors—equivalence and dependence of these divisors as primitive concepts. These notions will be related to each other by a set of simple axioms which, in the standard treatment of algebraic function fields of degree of transcendence one over algebraically-closed constant fields, are early theorems. We will show that these axioms characterize the divisor algebras of such fields.

1. Before beginning the study of our axiom system we should like to study briefly and state some of the results in a concrete case. We suppose that we are given an algebraic function field, K, in one variable, over an algebraically-closed constant field k.

A place P is a mapping of K into (k, ∞) such that:

1) $P(f) = 0$ if and only if $P(1/f) = \infty$.

2) If $P(f) \neq \infty$ and $P(g) \neq \infty$ then $P(f+g) = P(f) + P(g)$ and $P(fg) = P(f)P(g)$.

3) For every a in k, $P(a) = a$.

We will write $P(f)$ as $f(P)$. Let $\mathfrak{G}_K$ be the free Abelian group generated

---

* Received July 6, 1948.

[1] The material of this paper comes from a thesis, written under the direction of Professor Max A. Zorn, and submitted to the Graduate School of Indiana University for the degree of Doctor of Philosophy. The author wishes to take this opportunity to thank Professor Zorn for his guidance and encouragement.

[2] The numbers in square brackets refer to bibliography.

by the places. To every place P we can associate a valuation (logarithmic) V_P of K. We say a place P is a zero of f if $V_P(f) > 0$ and a pole of f if $V_P(f) < 0$. For only a finite number of places P is $V_P(f) \neq 0$. The divisor of a function f is defined to be $\prod_P P^{V_P(f)}$, and we say that $f \rightarrow \prod_P P^{V_P(f)}$. For any divisor $D = \prod_{i=1}^{n} P_i^{m_i}$ we define the order of D to be $\sum_{i=1}^{n} m_i$; if each $m_i \geq 0$, then D is said to be integral. We say that $A \mid B$ (read as A divides B) if $B = AC$ where C is an integral divisor. Two divisors A and B are defined to be equivalent if there exists an $f \varepsilon K$ with $f \rightarrow A/B$. It follows that this equivalence is reflexive, symmetric and transitive. We define $\mathfrak{H}$ to be the set of divisors equivalent to I, the unit element of $\mathfrak{G}_K$. $\mathfrak{H}$ is called the principal subgroup and its elements are called principal divisors. The equivalent divisors $A_1, \cdots, A_n$ are defined to be dependent if and only if the functions $f_i \rightarrow A_i/A_n$ are linearly dependent over k. By direct checking it can be ascertained that the divisors so constructed have the following properties:

1. $\mathfrak{H} \neq I$.

2. If A and B are equivalent, then they are of the same order.

3. The equivalence classes of divisors are projective geometries (in the sense of [2 pp. 208-209]) under the dependence.

4. If $A_1, \cdots, A_n$ are dependent, then for all divisors X, $A_1X, \cdots, A_nX$ are dependent.

5. A divisor dependent on integral divisors is itself integral.

6. If A and B are independent integral divisors and if P is any place, then there exists a divisor D_P, dependent on A and B, such that $P \mid D_P$.

7. Pappus' theorem is true in every plane of our geometries.

8. The transform of a projectivity by a divisor is again a projectivity.

If A, B, C are three independent points, we shall denote the plane generated by them as $[A, B, C]$, and the line generated by any two of them, say A and B, by $[A, B]$. By a marked line we shall mean a line from which three *distinct* points have been selected. If the three points selected are A, B, C then the marked line will be denoted by $\langle A, B, C \rangle$.

Following Hodge and Pedoe [2, chapter VI], on a marked line $\langle A, B, C \rangle$ we construct a field Φ_{ABC} plus an infinity. In lieu of property 7. above, these fields are all commutative. The elements of Φ_{ABC} will be the quadruples $(A,B,C;R)$ where $R\,\varepsilon\,\langle A,B,C\rangle$; and where $(A,B,C;A)$ is the zero element of Φ_{ABC}, $(A,B,C;B)$ is the unit element of Φ_{ABC} and $(A,B,C;C)$ is the infinity.

Let $f_1, f_2, f_3\,\varepsilon\,K$ be linearly independent over k and suppose that $f_1\rightarrow A$, $f_2\rightarrow C$ and $f_3\rightarrow D$. Thus A, C, D are independent divisors and so generate a plane $[A,C,D]$. Suppose further that $f_1+f_2\rightarrow B$. Consider the marked line $\langle A,B,C\rangle$ in the plane $[A,C,D]$. To every point X in $\langle A,B,C\rangle$ we can assign coordinates in the following two ways:

1) $(A,B,C;X)\,\varepsilon\,\Phi_{ABC}$.

2) If $\lambda f_1+\mu f_2\rightarrow X$ then the coordinates of X are $(\mu,0,\lambda)$ where λ, μ are in k. Thus, since for every non-zero $\rho\,\varepsilon\,k$

$$\rho(\lambda f_1+\mu f_2)\rightarrow X, \qquad \rho(\mu,0,\lambda)=(\mu,0,\lambda).$$

We define σ_{ABC} as follows: for $\lambda\neq 0$, $\sigma_{ABC}(\mu/\lambda)=(A,B,C;X)$. From Hodge-Pedoe [2, p. 270, paragraph 3] σ_{ABC} is an isomorphism of k onto Φ_{ABC}. Thus we have

THEOREM 1.1. *If K is an algebraic function field over an algebraically closed field of constants k, and if Φ_{ABC} is the field associated with the marked line $\langle A,B,C\rangle$ in a class of divisors of $\mathfrak{G}_K$, then $\sigma_{ABC}(k)=\Phi_{ABC}$ where σ_{ABC} is an isomorphism which can be given explicitly.*

Definition. If $f\rightarrow A/C$ and $f-1\rightarrow B/C$ then $f\leftrightarrow\langle A,B,C\rangle$.

A natural question would be: given two functions in K and the marked lines associated with them, what marked lines are associated with the sum and the product of these functions? The answer to these two questions is given by the following two theorems.

THEOREM 1.2. *If f, $g\,\varepsilon\,K$ are such that f, g, 1 are linearly independent over k, and if $f\leftrightarrow\langle A,B,C\rangle$ and $g\leftrightarrow\langle D,E,F\rangle$ then $f+g\leftrightarrow\langle R,S,CF\rangle$ where R and S are determined in the following manner* (Diagram 1).

Proof. Since f, g, 1 are linearly independent over k, CF, CD and AF are independent divisors. Since $S\,\varepsilon\,[CE,AF]\cap[CD,BF]$, $S/CF\,\varepsilon\,[E/F, A/C]\cap[D/F,B/C]$. Now $f\rightarrow A/C$, $f-1\rightarrow B/C$, $g\rightarrow D/F$ and $g-1\rightarrow E/F$; thus if $h\rightarrow S/CF$ then since $S/CF\,\varepsilon\,[A/C,E/F]$, $h=\lambda(g-1)+\mu f$ where λ, μ are in k. But since S/CF is in $[D/F,B/C]$, $h=\alpha g+\beta(f-1)$ where $\alpha,\beta\,\varepsilon\,k$. Since $f, g, 1$ are linearly independent over k, $\alpha=\beta=\lambda=\mu$, and hence $f + g - 1 \to S/CF$. Using the same argument for R/CF we obtain that $f + g \to R/CF$. Combining the two we have that $f + g \leftrightarrow \langle R, S, CF \rangle$.

If f, g, 1 are dependent linearly over k, and $f \neq -g + \mu$, then if $g = \lambda f + \mu$ it follows from Theorem 1. 1 that $\sigma_{ABC}(-\mu/\lambda) = (A, B, C; R)$ and that $\sigma_{ABC}(-\mu/\lambda + 1) = (A, B, C; S)$ implies that $g \leftrightarrow \langle R, S, C \rangle$. We then find that $f + g \leftrightarrow \langle M, N, C \rangle$ where $\sigma_{ABC}(-\mu/(\lambda + 1) = (A, B, C; M)$ and $\sigma_{ABC}(-\mu/(\lambda + 1) + 1) = (A, B, C; N)$.

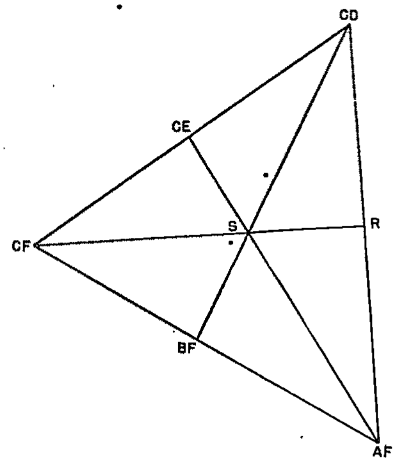


DIAGRAM 1.

Using the same method of proof as in Theorem 1. 2 we readily obtain

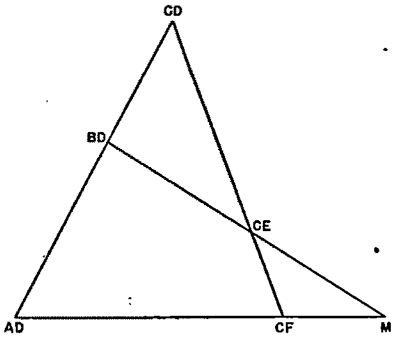


DIAGRAM 2.

THEOREM 1. 3. *If* $f, g \in K$ *and* $fg \notin k$ *and if* $f \leftrightarrow \langle A, B, C \rangle$ *and* $g \leftrightarrow \langle D, E, F \rangle$ *then* $fg \leftrightarrow \langle AD, M, CF \rangle$ *where* M *is determined in the following manner* (Diagram 2).

We shall return to these considerations again at the end of this paper.

2. The primitive concepts and axioms. In this section the primitive concepts are to be introduced and an axiom system set up to relate these fundamental concepts and the objects defined from them.

I. The first primitive concept to be considered is that of prime divisor. We assume that we are given a set of objects which we will call prime divisors and will denote by P, P_i, Q, Q_i etc. The set of prime divisors will be called the Riemann surface, $\mathfrak{R}$.

Let $\mathfrak{G}$ be the free Abelian group generated by the prime divisors. We identify P with P^1. Thus $\mathfrak{R} \subset \mathfrak{G}$. By a divisor D we mean an element of $\mathfrak{G}$. Thus the unit divisor I. of $\mathfrak{G}$, is $\prod_{P \varepsilon \mathfrak{R}} P^0$.

Definition. A divisor $D = \prod_{i=1}^{n} P_i^{\alpha_i}$ is said to be an integral divisor if for each $i = 1, 2, \cdots, n$, $\alpha_i \geq 0$.

Since $\mathfrak{G}$ is Abelian there will be no confusion in writing AB^{-1} as A/B. It is clear that every divisor D can be written as A/B where A and B are integral divisors.

Definition. If A and B are divisors then $A \mid B$ (read as A divides B) if and only if $B = AC$ where C is an integral divisor.

Two integral divisors A and B will be said to be relatively prime if for all prime divisors P the statement $P \mid A$ and $P \mid B$ is false.

Definition. If $D = \prod_{i=1}^{n} P_i^{\alpha_i}$ then the order of D, denoted by $n(D)$, is $\sum_{i=1}^{n} \alpha_i$. Thus $n(I) = 0$ where I is the unit divisor.

From the definitions of the order of a divisor and the product of two divisors we immediately have

LEMMA 1. *For all A, $B \varepsilon \mathfrak{G}$, $n(AB) = n(A) + n(B)$ and $n(A/B) = n(A) - n(B)$.*

It is clear that if $n(D) < 0$ then D is not integral, and if $n(D) = 0$ then D is integral if and only if $D = I$.

II. The next primitive concept that we introduce is that of the principal subgroup $\mathfrak{H}$ of $\mathfrak{G}$. A divisor which is in $\mathfrak{H}$ will be termed a principal divisor.

Definition. If A and B are divisors then $A \sim B$ if and only if A/B is in $\mathfrak{H}$.

It is trivial that if $A \sim B$ then for all divisors X, $AX \sim BX$. We will denote the equivalence class of a divisor A by (A). The set of integral divisors of (A) will be denoted by $\{A\}$.

III. The third and last primitive concept that will concern us is that of dependence relations defined only for equivalent divisors. We suppose that, given $A_1, \cdots, A_n$ as equivalent divisors we can say whether or not they are dependent. If the A_i are not dependent they will be said to be independent. If the A_i are not equivalent, neither dependence nor independence will be defined for them. If $A_1, \cdots, A_n$, B are dependent and if for some s, $0 < s \leq n$, $A_1, \cdots, A_s$, B are dependent, while $A_1, \cdots, A_s$ are independent, then B is said to depend on $A_1, \cdots, A_n$.

Now that our primitive concepts have been laid out and some definitions derived from them, we introduce a system of axioms to intertwine these concepts and definitions.

The axioms are as follows:

1. $\mathfrak{H} \neq I$.

2. If $A \sim B$ then $n(A) = n(B)$.

3. Axioms of dependence:

a) If A depends on $B_1, \cdots, B_{n-1}, B_n$ but not on $B_1, \cdots, B_{n-1}$, then B_n depends on $B_1, \cdots, B_{n-1}, A$.

b) If A depends on $B_1, \cdots, B_n$ and each B_i depends on $C_1, \cdots, C_m$, then A depends on $C_1, \cdots, C_m$.

c) If $A_1, \cdots, A_n$ are independent and $B_1, \cdots, B_m$ are independent, but $A_1, \cdots, A_n, B_1, \cdots, B_m$ are dependent, then there exists a divisor D which depends on $A_1, \cdots, A_n$ and on $B_1, \cdots, B_m$.

d) A and B are dependent if and only if $A = B$.

4. If $A_1, \cdots, A_n$ are dependent, then for all divisors X, $A_1X, \cdots, A_nX$ are dependent.

5. If B depends on $A_1, \cdots, A_n$ which are all integral divisors, then B is integral.

6. If A and B are two independent integral divisors, and if P is any prime divisor, then there exists a divisor D_P dependent on A and B and such that $P \mid D_P$.

The following axiom contains a concept which as yet is not defined.

It will be applied only after the concept of multiprojectivity has been formally defined.

7. The identity multiprojectivity is the only multiprojectivity leaving three distinct, dependent points fixed.[3]

3. Linear families.

In this section we will study some of the properties of the dependence relation and some of the immediate consequences of the axioms in regard to these relations.

If $A_1, \cdots, A_n$ is a sequence of equivalent divisors, then by the linear family $[A_1, \cdots, A_n]$ we mean the set of divisors B which depend on $A_1, \cdots, A_n$. If s is the maximum number of independent A_i, $i = 1, \cdots, n$, then s will be called the dimension of $[A_1, \cdots, A_n]$ and denoted by $\dim [A_1, \cdots, A_n]$. From the abstract theory of linear dependence [3, vol. 1, p. 95] the dimension of a linear family is invariant and any subfamily of a finite dimensional family is itself finite dimensional. By the multiples of a divisor D in a linear family F we mean the set of divisors X in F such that $D \mid X$.

THEOREM 3.1. *If $A_1, \cdots, A_n$ are equivalent divisors and for each $i = 1, 2, \cdots, n$ $D \mid A_i$, then $D \mid X$ for all X in $[A_1, \cdots, A_n]$.*

Proof. X depends on $A_1, \cdots, A_n$, hence by axiom 4 X/D depends on $A_1/D, \cdots, A_n/D$. But since $D \mid A_i$ for $i = 1, \cdots, n$, the A_i/D are integral and thus by axiom 5 X/D is integral. Thus $D \mid X$.

COROLLARY 1. *The multiples of D in $[A_1, \cdots, A_n]$ form a linear family.*

COROLLARY 2. *If A and B are relatively prime, independent integral divisors, then for each prime divisor P there exists exactly one divisor D_P in $[A, B]$ such that $P \mid D_P$.*

Proof. By axiom 6 there exists at least one such D_P. Suppose that there is another one $D'_P \neq D_P$. Then by axioms 3 a. and 3 b. $[A, B] = [D_P, D'_P]$. Thus by Theorem 3.1 P/A and $P \mid B$ which contradicts that A is relatively prime to B.

---

[3] This axiom could be replaced by the following two axioms which are more easily verified in the case of a concrete function field:

a. Pappus' Theorem.

b. The transform of a projectivity by a divisor is again a projectivity.

THEOREM 3.2. *If* $A_1, \cdots, A_n$ *are equivalent integral divisors and* $\dim [A_1, \cdots, A_n] = s$, *then for every prime divisor* P *the multiples of* P *in* $[A_1, \cdots, A_n]$ *form a linear family of dimension at least* $s - 1$.

Proof. If $P \mid A_i$ for $i = 1, \cdots, n$ then the multiples of P in $[A_1, \cdots, A_n]$ by Theorem 3.1 are exactly $[A_1, \cdots, A_n]$. Suppose then that $P \nmid A_1$. Let X be any divisor in $[A_1, \cdots, A_n]$ such that $X \neq A_1$ and $P \nmid X$. By axiom 6 there exists an X_P in $[X, A_1]$ such that $P \mid X_P$. Since X_P depends on X and A_1 and is different from both of them, X depends on A_1 and X_p. Thus the multiples of P in the family $[A_1, \cdots, A_n]$ together with A_1 generate all of $[A_1, \cdots, A_n]$; hence they form an $s - 1$ dimensional family.

THEOREM 3.3. *For every divisor* A, $\dim \{A\}$ *is finite.*

Proof. By axiom 5, $\{A\}$ is closed under dependence.

By induction over $n(A)$ we have:

1) If $n(A) < 0$ then $\{A\}$ is the null set and so the theorem is true.

2) Suppose the theorem is true for all divisors of order less than r. Let $n(A) = r$. If P is any prime divisor then $n(A/P) = r - 1$ and so by our induction hypothesis $\dim \{A/P\} = s$, say, where s is finite. Thus $\{A/P\} = [B_1, \cdots, B_s]$. Suppose that a C and a D could be found in $\{A\}$ such that $B_1P, \cdots, B_nP, C, D$ are independent. Then by Theorem 3.2 the multiples of P in $[B_1P, \cdots, B_nP, C, D]$ form a linear family of dimension at least $s + 1$. But this implies that $\dim \{A/P\} = s \geq s + 1$, which is false. Thus $\dim \{A\} \leq s + 1$. This concludes the induction.

THEOREM 3.4. *If* $A \mid B$ *then* $n(A) - \dim \{A\} \leq n(B) - \dim \{B\}$.

Proof. In the course of the proof of Theorem 3.3 we showed that for all prime divisors P, $\dim \{AP\} \leq \dim \{A\} + 1$. Thus if $B = AC$ where C is an integral divisor, then applying the result quoted $n(C)$ times we obtain the theorem.

THEOREM 3.5. *There exists an* A *such that* $\dim \{A\} \geq 2$.

Proof. By axiom 1, $\mathfrak{H} \neq I$. Hence there exists an $R \in \mathfrak{H}$, $R \neq I$. Now $R = A/B$ where A and B are distinct integral divisors; since $R \in \mathfrak{H}$, $A \sim B$. Thus $\dim \{A\} \geq 2$ from axiom 3 d.

THEOREM 3.6. *If* A *and* B *are independent integral divisors then for all positive integers* n, $A^n, A^{n-1}B, \cdots, AB^{n-1}, B^n$ *are independent divisors.*

Proof. Without loss of generality A and B can be taken as relatively prime. For if $A = CX$ and $B = DX$ where C and D are relatively prime, then to check the independence of $A^n, A^{n-1}B, \cdots, AB^{n-1}, B^n$, that is of $C^nX^n, C^{n-1}X^nD, \cdots, CD^{n-1}X^n, D^nX^n$ it is sufficient, by axiom 4, to check the independence of $C^n, C^{n-1}D, \cdots, CD^{n-1}, D^n$.

By induction over n we have:

1) If $n = 1$, then by hypothesis A^1, B^1 are independent.

2) Suppose the theorem is true for $n = r \geqq 1$.

Consider $A^{r+1}, A^rB, \cdots, AB^r, B^{r+1}$. By the induction hypothesis, A^r, $A^{r-1}B, \cdots, AB^{r-1}, B^r$ are independent, hence by axiom 4, $A^{r+1}, A^rB, \cdots, AB^r$ are independent. So if $A^{r+1}, A^rB, \cdots, AB^r, B^{r+1}$ are dependent, then B^{r+1} depends on $A^{r+1}, A^rB, \cdots, AB^r$. Therefore from Theorem 3.1, $A \mid B^{r+1}$, which is impossible since A and B are relatively prime and $A \neq I$.

Definition. If A and B are independent divisors, then $[A, B]^n = [A^n, A^{n-1}B, \cdots, AB^{n-1}, B^n]$.

COROLLARY. *Dim* $[A, B]^n = n + 1$.

THEOREM 3.7. *For every integer n there exists a divisor A_n such that dim $\{A_n\} > n$.*

Proof. By Theorem 3.5 there exists an A such that dim $\{A\} \geqq 2$. Let A and B be two independent divisors in $\{A\}$. Then $[A, B]^n \subset \{A^n\}$. But dim $[A, B]^n = n + 1$, hence dim $\{A^n\} \geqq n + 1$. Let $A_n = A^n$.

4. Linear families as projective geometries. We are now going to consider the integral divisors of a class, (any finite dimensional linear family will do) and show that by the proper definition of points, lines, etc. the integral divisors of a class form a finite dimensional, Desarguian projective space. As axioms for projective geometry we are using those of Hodge and Pedoe [2, pp. 208-209]. It should be noted here that the dimension of a linear family that we use here is always one larger than that of Hodge-Pedoe.

THEOREM 4.1. *The integral divisors of a class form a finite dimensional, Desarguian projective space.*

Proof. We take the integral divisors of the class to be points of our space, and for the Hodge-Pedoe linear spaces S_h we take the linear families $[A_1, \cdots, A_h]$ where the A_i are independent integral divisors of the class.

By direct checking it is easily seen that these satisfy the Hodge-Pedoe axioms for projective geometries. Theorem 3.3 gives us the finite dimensionality. There remains but to show the Desargues theorem.

If $\dim\{A\} > 3$ the space is Desarguian. If $\dim\{A\} = 3$ then by multiplying by an appropriate X we can, by Theorem 3.7, obtain a divisor AX such that $\dim\{AX\} > 3$. Since multiplication by X preserves dependence, it preserves collinearity in the projective space; thus the Desargues theorem for $\{AX\}$ implies the Desargues theorem for $\{A\}$.

By a marked line we shall mean a linear family $[X, Y]$ from which three *distinct* points A, B, C have been selected. We shall denote the line marked by A, B, C by $\langle A, B, C\rangle$.

Following Hodge-Pedoe [2, chapter VI], on a marked line $\langle A, B, C\rangle$ we construct a field Φ_{ABC} plus an infinity. The elements of Φ_{ABC} will be quadruples $(A, B, C; R)$ where R is in $\langle A, B, C\rangle$; and where $(A, B, C; A)$ is the zero element of Φ_{ABC}, $(A, B, C; B)$ is the unit element of Φ_{ABC} and $(A, B, C; C)$ is the infinity.

Given a sequence of divisors which are independent, then we can use the hyperplanes $\eta_i = [A_1, \cdots, A_{i+1}, A_{i+1}, \cdots, A_n]$ (where n is the dimension of the space) as coordinate planes, as is done in Hodge-Pedoe [2, p. 262]. Thus we introduce an analytic projective geometry where the equation of any hyperplane is linear. Whenever the need arises we shall feel free to use analytic projective geometry.

By a perspectivity σ of one line α onto another line β from a point $R \notin \alpha$, $R \notin \beta$, we shall mean the mapping defined for every X by $\sigma(X) = [X, R] \cap \beta$ where $[X, R] \cap \beta$ is the point of intersection of $[X, R]$ and β, A projectivity is the product of a finite number of perspectivities. We will denote projectivities by τ.

σ_X is defined for all A by $\sigma_X(A) = XA$. Very often we will be considering the resulting image of a line under a σ_X and in this case we will consider σ_X as a line-to-line mapping and will not use a different notation for it.

Definition. A multiprojectivity is a mapping which maps a line in a space A onto a line in a second space B, which can be written as the product of a finite number of τ's and σ_X's.

LEMMA 2. *If π_1 and π_2 are multiprojectives of $\langle A, B, C\rangle$ onto $\langle D, E, F\rangle$ such that $\pi_1(A) = \pi_2(A) = D$, $\pi_1(B) = \pi_2(B) = E$ and $\pi_1(C) = \pi_2(C) = F$ then $\pi_1 = \pi_2$.*

Proof. $\pi_2^{-1}\pi_1$ is a multiprojectivity and it leaves A, B, C respectively, fixed. Hence from axiom 7. it is the identity mapping; hence the lemma.

As a special case of this lemma if τ is a projectivity taking three collinear points into three other specified collinear points, then it is unique. This implies Pappus' theorem in every plane of our geometries, which in turn implies the commutativity of the fields Φ_{ABC} for any $\langle A, B, C\rangle$.

For completeness we will now state a well-known projective theorem.

THEOREM 4. 2. *If τ is a projectivity of $\langle A, B, C\rangle$ onto $\langle R, S, T\rangle$ which takes A into R, B into S and C into T, then the mapping τ^* defined by $\tau^*(A, B, C; D) = R, S, T; \tau(D))$ is an isomorphism of Φ_{ABC} onto Φ_{RST}.*

We are naturally concerned with the relationship between Φ_{ABC} and $\Phi_{AX,BX,CX}$ for any given X. Using an argument which is of the same nature as that used in proving Theorem 4. 2 we have

THEOREM 4. 3. *If $\sigma_X : \langle A, B, C\rangle \rightarrow \langle AX, BX, CX\rangle$ such that $\sigma_X(R) = XR$ then the mapping σ^*_X defined by $\sigma^*_X(A, B, C; R) = (AX, BX, CX; RX)$ is an isomorphism of Φ_{ABC} onto $\Phi_{AX,BX,CX}$.*

Combining Theorems 4. 2 and 4. 3 we immediately have

THEOREM 4. 4. *If $\pi_{ABC}{}^{RST}$ is a multiprojectivity of $\langle A, B, C\rangle$ onto $\langle R, S, T\rangle$ which takes A into R, B into S and C into T, then the mapping $\pi^*_{ABC}{}^{RST}$ defined by $\pi^*_{ABC}{}^{RST}(A, B, C; D) = (R, S, T; \pi_{ABC}{}^{RST}(D))$ is an isomorphism of Φ_{ABC} onto Φ_{RST}.*

Given two marked lines $\langle A, B, C\rangle$ and $\langle R, S, T\rangle$ then the mapping $\sigma(X) = XR/A$ brings the first line into the plane of the second. Then from projective geometry we can find a projectivity that will take BR/A into S, CR/A into T and R into itself. Thus we can find a multiprojectivity which will take one marked line onto any other. Thus, from Theorem 4. 4 we have that the fields constructed on any two marked lines are isomorphic, and that an isomorphism can be explicitly exhibited.

5. The function field.

THEOREM 5. 1. *If A and B are independent integral divisors, then for every prime divisor P there exists an integer $n > 0$ and a divisor D_P in $[A, B]$ such that $P^n | D_P$ and if $P^n | X$ in $[A, B]$, then $X = D_P$.*

Proof. Let $A = A'S$ and $B = B'S$ where S is an integral divisor and A' and B' are relatively prime integral divisors. Thus $[A, B] = [A'S, B'S]$

$= S[A', B']$. Applying Corollary 2, Theorem 3.1 to $[A', B']$ we obtain a unique D'_P in $[A', B']$, such that $P \mid D'_P$. Let $D_P = D'_P S$.

Thus to every line of integral divisors, $[A, B]$, and to every prime divisor P we can associate a unique D_P in $[A, B]$. This gives us a function mapping the Riemann surface $\mathfrak{R}$ onto any line $[A, B]$ of integral divisors. As will be indicated, the mapping can easily be extended to any line $[C, D]$. We define our mapping precisely and study the nature of the functions so obtained.

Definition. If A and B are independent integral divisors, then $F_{AB}: \mathfrak{R} \to [A, B]$ where for every $P \varepsilon \mathfrak{R}$, $F_{AB}(P) = D_P \varepsilon [A, B]$ and D_P is as defined in Theorem 5.1.

If A and B are independent, but not integral, divisors, then we multiply them by S so that AS and BS are integral. We define F_{AB} by $F_{AB}(P) = S^{-1} F_{AS,BS}(P)$. From Theorem 3.1 it follows that this definition is independent of S.

If we are given two marked lines we have pointed out that the fields constructed on them are isomorphic and that an isomorphism can be explicitly given. We pick a particular marked line $\langle X, Y, Z \rangle$ and the field $\Phi_{XYZ} = \Phi$ and consider them as fixed henceforth. Given $\langle A, B, C \rangle$, by the multiprojectivity π_{ABC} we mean that multiprojectivity of $\langle A, B, C \rangle$ onto $\langle X, Y, Z \rangle$ which takes A into X, B into Y and C into Z. To a marked line $\langle A, B, C \rangle$ we now associate a function f_{ABC} as follows:

Definition. $f_{ABC}: \mathfrak{R} \to \Phi$ where for every $P \varepsilon \mathfrak{R}$, $f_{ABC}(P) = (X, Y, Z; \pi_{ABC} F_{AB}(P))$.

We define two such functions f_{ABC} and f_{RST} to be equal if for all P such that $f_{ABC}(P) \neq (X, Y, Z; Z) \neq f_{RST}(P)$, $f_{ABC}(P) = f_{RST}(P)$.

THEOREM 5.2. *For all divisors R, $f_{ABC} = f_{AR,BR,CR}$.*

Proof. From the definition of $F_{AR,BR}$ and Theorem 3.1 we have that for all P, $F_{AR,BR}(P) = RF_{AB}(P)$. Let $\sigma_R(D) = RD$ for all divisors D. Thus $\pi_{AR,BR,CR}\sigma_R$ is a multiprojectivity and it takes A into X, B into Y and C into Z. Thus by Lemma 2, it follows that $\pi_{ABC} = \pi_{AR,BR,CR}\sigma_R$. Thus for every P,

$$\begin{aligned} f_{AR,BR,CR}(P) &= (X, Y, Z; \pi_{AR,BR,CR} F_{AR,BR}(P)) \\ &= (X, Y, Z; \pi_{ABC}\sigma^{-1}{}_R R F_{AB}(P)) \\ &= (X, Y, Z; \pi_{ABC} F_{AB}(P)) = f_{ABC}(P). \end{aligned}$$

We note here that because of the isomorphism of Φ_{ABC} onto Φ by π^*_{ABC} it is sufficient to carry out any calculation in Φ_{ABC} and derive the result for Φ using this isomorphism.

Now that a function has been assigned to every marked line, we attempt to define compositions for these functions so that under these compositions they form a [illegible] We begin with the sum of two functions.

If [illegible] $= (X, Y, Z; Z)$ we will say that $f_{ABC}(P) = \infty$.

Definition. If f_{ABC} and f_{DEF} are two functions such that CF, CD and AF are independent divisors, then $f_{ABC} + f_{DEF} = f_{R,S,CF}$ where R and S are determined as in Diagram 1.

From its definition, the addition of functions is commutative. We now proceed to show that $f_{R,S,CF}$ is equal to the sum of the two functions in our sense.

THEOREM 5.3. *If f_{ABC} and f_{DEF} satisfy the conditions of the preceding definition, then for all P such that $f_{ABC}(P) \neq \infty$ and $f_{DEF}(P) \neq \infty$, $(f_{ABC} + f_{DEF})(P) = f_{ABC}(P) + f_{DEF}(P)$.*

Proof. Consider Diagram 1. As was pointed out before, we can assign coordinates, coming from the field $\Phi_{CD,R,AF}$, to the plane and use analytic projective geometry. Lower case Greek letters will indicate elements of $\Phi_{CD,R,AF}$. We assign coordinates axes as follows:

$[CF, AF]$ as the x-axis, $[CF, CD]$ as the y-axis and $[CD, AF]$ as the z axis. We assign, as coordinates, to CF, $(0,0,1)$; to CE, $(1,0,1)$; to CD, $(1,0,0)$; to BF, $(0,1,1)$; to AF, $(0,1,0)$ and to R, $(1,1,0)$. We map $\langle CF, CE, CD\rangle$ onto $\langle CD, R, AF\rangle$ by the perspectivity σ_2 defined by $\sigma_2(\lambda, 0, \mu) = (\lambda, \mu, 0)$. Evidently CD goes into CD, CE into R and CF into AF. We map $\langle AF, BF, CF\rangle$ onto $\langle CD, R, AF\rangle$ by the projectivity σ_1, defined by $\sigma_1(0, \alpha, \beta) = (\alpha, \beta, 0)$. Evidently σ_1 takes AF into CD, BF into R and CF into AF. Since the equation of $[CF, R]$ is $x = y$, the coordinates of a point on it are of the form (γ, γ, δ). We map $\langle CF, S, R\rangle$ onto $\langle CD, R, AF\rangle$ by the projectivity σ_3 defined by $\sigma_3(\gamma, \gamma, \delta) = (\gamma, \delta, 0)$. Evidently CF is taken into AF, S into R and R into CD by σ_3.

If $(\lambda, 0, \mu)$, $(0, \alpha, \beta)$ and (γ, γ, δ) are collinear, then

$$\begin{vmatrix} \gamma & \gamma & \delta \\ \lambda & 0 & \mu \\ 0 & \alpha & \beta \end{vmatrix} = -\gamma(\alpha\mu + \beta\lambda) + \delta\lambda\alpha = 0.$$

Thus if $\alpha \neq 0$ and $\lambda \neq 0$, $\delta/\gamma = \beta/\alpha + \mu/\lambda$.

By Theorem 5.2, given f_{ABC} and f_{DEF}, we may always assume that A and C are relatively prime integral divisors; and we may also make this assumption in regards to D and F. Thus $f_{ABC}(P) = \infty$
similarly $f_{DEF}(P) = \infty$ if and only if $P \mid F$. From Theor
of any prime divisor P in the family $[CF, CD, AF]$ are
sional family. If the multiples of P are all of the plane,
and $P \mid AF$. Hence by the relative primeness of A an
$P \mid C$ and $P \mid F$. Hence if $f_{ABC}(P) \neq \infty$ and $f_{DEF}(P) \neq$
P in the plane form a line. Thus $F_{RS}(P)$, $F_{AF,CF}(P)$
collinear. Hence if their coordinates are (γ, γ, δ), $(0, \alpha, \beta)$ and $(\lambda, 0, \mu)$, respectively, then $\alpha \neq 0$ and $\lambda \neq 0$ and $\delta/\gamma = \beta/\alpha + \mu/\lambda$. σ_1, σ_2, σ_3, are the multiprojectivities used for defining f_{ABC}, f_{DEF} and $f_{R,S,CF}$ relative to $\Phi_{CD,R,AF}$ and $f_{ABC}(P) = \beta/\alpha$, $f_{DEF}(P) = \mu/\lambda$ and $f_{R,S,CF}(P) = \delta/\gamma$. Hence the theorem.

Definition. If $\lambda \neq 0 \in \Phi$ then $\lambda f_{ABC} = f_{AB'C}$ where $\lambda^{-1} = (X, Y, Z; \pi_{ABC}(B'))$.

THEOREM 5.4. *For every prime divisor P, $(\lambda f_{ABC})(P) = \lambda(f_{ABC}(P))$.*

Proof. Let $\lambda f_{ABC} = f_{AB'C}$. Since $B' \neq A$, $\pi_{ABC}(B') = R \neq X$. We define σ as follows:

$(X, Y, Z; \sigma(T)) \cdot (X, Y, Z; R) = (X, Y, Z; T)$. Thus σ is a projectivity, and $\sigma\pi_{ABC}$ is a multiprojectivity. Since $(X, Y, Z; X) \cdot (X, Y, Z; R) = (X, Y, Z; X)$, $\sigma\pi_{ABC}(A) = \sigma(X) = X$. Similarly $\sigma\pi_{ABC}(C) = Z$. Since $(X, Y, Z; \sigma(R)) \cdot (X, Y, Z; R) = (X, Y, Z; R)$, $\sigma(R) = Y$. Thus $\sigma\pi_{ABC}(B') = Y$, from the definition of R. But $\pi_{AB'C}$ is also a multiprojectivity taking A into X, B' into Y and C into Z. Thus from Lemma 2, $\sigma\pi_{ABC} = \pi_{AB'C}$. Thus we have,

$$\begin{aligned}(\lambda f_{ABC})(P) &= f_{AB'C}(P) = (X, Y, Z; \pi_{AB'C}F_{AC}(P)) = (X, Y, Z; \sigma\pi_{ABC}F_{AC}(P)) \\ &= (X, Y, Z; \pi_{ABC}F_{AC}(P)) \cdot (X, Y, Z; R)^{-1} \text{ from the definition of } \sigma, \\ &= f_{ABC}(P)\lambda \text{ since } \lambda = (X, Y, Z; R)^{-1}.\end{aligned}$$

Hence we have the theorem.

Definition. If $\lambda \in \Phi$ then $f_{ABC} + \lambda = f_{A'B'C}$ where $-\lambda = (X, Y, Z; \pi_{ABC}(A'))$, and $-\lambda + 1 = (X, Y, Z; \pi_{ABC}(B'))$.

Using the exact type of argument as in the proof of Theorem 5.4, with the obvious modifications we obtain

THEOREM 5.5. *For every prime divisor P,*

$$(f_{ABC}+\lambda)(P)=f_{ABC}(P)+\lambda.$$

s, essentially the converses of Theorems 5.4 and
g these lines.

ery prime divisor P, $f_{AB'C}(P)=\lambda^{-1}f_{ABC}(P)$ *where*
$\neq 0$.

THEOREM 5.7. *If A, A', and C are dependent divisors, then* $f_{ABC}=\lambda f_{A'B'C}+\mu$ *where* $\lambda,\mu\ \varepsilon\ \Phi$.

We now return to the case of the sum of f_{ABC} and f_{DEF} where CF, AF and CD are dependent. Since we may assume that A and C, and D and F are relatively prime integral divisors (Theorem 5.2), from A and C prime to each other it follows by Theorem 3.1 that $C\,|\,F$. Similarly we can obtain that $F\,|\,C$. Hence $F=C$. Thus A, D, C are dependent, and so from Theorem 5.7 we have that $f_{ABC}=\lambda f_{DEF}+\mu$ where $\lambda,\mu\ \varepsilon\ \Phi$. If $\lambda\neq-1$, we define $f_{ABC}+f_{DEF}=f_{RSC}$ where $f_{RSC}=(\lambda+1)f_{DEF}+\mu$. Applying Theorems 5.5 and 5.6 we have that in this case also that for those prime divisors P for which neither function is infinite, $(f_{ABC}+f_{DEF})(P)=f_{ABC}(P)+f_{DEF}(P)$. If $\lambda=-1$, we define $f_{ABC}+f_{DEF}=\mu$ and again for those prime divisors P for which neither function is infinite, $(f_{ABC}+f_{DEF})(P)=f_{ABC}(P)+f_{DEF}(P)$.

Now that addition has been defined and shown to have some desirable properties, we turn to multiplication. We proceed as follows.

Definition. If f_{ABC} and f_{DEF} are such that AD, CD, and CF are independent divisors, then $f_{ABC}f_{DEF}=f_{AD,M,CF}$ where M is determined as in Diagram 2.

Using analytic projective geometry, and proceeding as in the proof of Theorem 5.3 we obtain

THEOREM 5.8. *If* f_{ABC} *and* f_{DEF} *satisfy the conditions of the preceding definition, then for all P such that* $f_{ABC}(P)\neq\infty$, *and* $f_{DEF}(P)\neq\infty$,

$$(f_{ABC}f_{DEF})(P)=f_{ABC}(P)f_{DEF}(P).$$

We now return to the definition of multiplication. Suppose that f_{ABC} and f_{DEF} are two functions. Without loss of generality we may assume that A and C, and also D and F are relatively prime integral divisors. If AD, CD, and CF are independent, then by definition $f_{ABC}f_{DEF}=f_{AD,M,CF}$; and if AD, AF and CF are independent, $f_{DEF}f_{ABC}=f_{AD,N,CF}$ (see Diagram 3).

Pick a P so that $P \nmid CF$ and $P \nmid AD$. Then there is
that $P | D_P$. Since $f_{ABC}(P) \neq \infty$ and $f_{DEF}(P) \neq \infty$,

$$(f_{ABC} f_{DEF})(P) = f_{ABC}(P) f_{DEF}(P) = f_{DEF}(P) f_{ABC}($$

Thus $\pi_{AD,N,CF} D_P = \pi_{AD,M,CF} D_P$. Since the two mult
on the distinct points AD, CF and D_P, the multiproje
so $M = N$, and the multiplication is commutative, in

If AD, CD, and CF are dependent, then $C | D$ a
If in addition AF, AD, and CF are also dependent, t
this case $f_{DEF} = f_{CEA}$. Combining the following theorem and Theorem 5.6, then $f_{DEF} = \lambda f_{ABC}^{-1}$ and for all P for which f_{DEF} is not infinite, $f_{ABC}(P) f_{DEF}(P) = \lambda$. We define $f_{ABC} f_{DEF} = \lambda$.

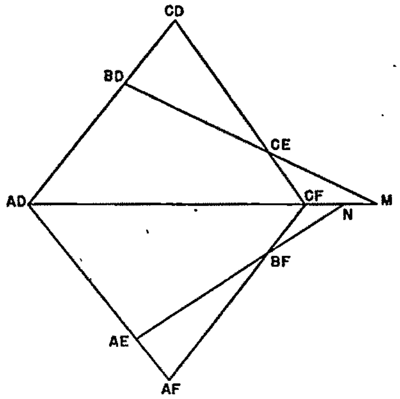


DIAGRAM 3.

If AF, AD, and CF are independent then $f_{DEF} f_{ABC}$ can be constructed. In this case we define $f_{ABC} f_{DEF} = f_{DEF} f_{ABC}$. Thus for all P such that $f_{ABC}(P) \neq \infty$ and $f_{DEF}(P) \neq \infty$ we have

$$f_{ABC} f_{DEF})(P) = (f_{DEF} f_{ABC})(P) = f_{DEF}(P) f_{ABC}(P) = f_{ABC}(P) f_{DEF}(P).$$

Hence in all cases $(f_{ABC} f_{DEF})(P) = f_{ABC}(P) f_{DEF}(P)$; moreover the multiplication is commutative.

THEOREM 5.9. *For all P such that $f_{ABC}(P) \neq \infty$, $f_{ABC}(P) = (f_{CBA}(P))^{-1}$.*

Proof. We define σ as follows:

(1) $\sigma(X) = Z$;

(2) $\sigma(Z) = X$;

id $R \neq Z$, $(X, Y, Z; \sigma(R)) = (X, Y, Z; R)^{-1}$.
σ is a projectivity, $\sigma\pi_{ABC}$ is a multiprojectivity;
f of Theorem 5.4 we obtain this theorem.

an the marked line $\langle RG, SG, TG\rangle$.

$f_{ABC} = f_{RST}$ then for some divisor G, $\langle A, B, C\rangle$

and CR are independent, then $f_{ABC}f_{TSR}$ is not a
$_{TSR} = 1$ from Theorem 5.9, CT, AT, and CR are
dependent. Without loss of generality we can assume that A and C and also T and R are relatively prime integral divisors. Then $C = T$ since $C \mid T$ and $T \mid C$. Thus C, A and R are dependent. Since A and C are relatively prime and since $R \in [A, C]$, $R \neq A$ would imply that there exists a P with $P \mid A$ and $P \nmid R$. Thus $f_{ABC}(P) = (X, Y, Z; X)$ and $f_{RST}(P) \neq (X, Y, Z; X)$, contradicting the hypothesis. Thus $A = R$. Similarly $B = S$. If A and C are not relatively prime, then writing $A = A'U$ and $C = C'U$, where A' and C' are relatively prime and integral, the result comes immediately. Combining Theorems 5.2 and 5.10 we have

COROLLARY. *$f_{ABC} = f_{RST}$ if and only if $\langle A, B, C\rangle = G\langle R, S, T\rangle$ for some G.*

LEMMA 3. *If $f_{ABC}(P) = f_{DEF}(P)$ for all but a finite number of P's, then $f_{ABC} = f_{DEF}$.*

Proof. Without loss of generality, A and C, and also D and F, are relatively prime integral divisors. Since, for all but a finite number of P $(f_{ABC}f_{FED})(P) = f_{ABC}(P)f_{FED}(P) = f_{ABC}(P)(f_{DEF}(P))^{-1} = 1$, (hypothesis and Theorem 5.9), $f_{ABC}f_{FED}$ takes on only a finite number of values. But this can only happen when $AF = CD$. From the relative primeness of the factors involved, $A = D$ and $F = C$. From the uniqueness of a multiprojectivity taking three collinear point into three others, it follows that $B = E$. Thus the lemma.

Using the results of Theorems 5.3-5.10 and the above lemma, we easily obtain

THEOREM 5.11. *For all functions f_{ABC}, f_{DEF}, f_{GHK}, we have*

1. $(f_{ABC} + f_{DEF}) + f_{GHK} = f_{ABC} + (f_{DEF} + f_{GHK})$;

2. $(f_{ABC}f_{DEF})f_{GHK} = f_{ABC}(f_{DEF}f_{GHK})$;

3. $f_{ABC}(f_{DEF}+f_{GHK})=f_{ABC}f_{DEF}+f_{ABC}f_{GHK}$.

Let K_G be the set of all functions f_{ABC} (that is, marked line) with the field Φ adjoined. Thus we c operations which are commutative, associative, distrik Theorem 5. 6) we have a multiplicative unit, and additive unit element. Theorem 5. 9 gives us the mu if we define $-f=(-1)f$ we obtain

THEOREM 5. 12. *K_G is a field under the compos* [illegible]

6. Linear dependence in K_G. In this section we will study the relation between the dependence of a finite sequence of principal divisors, (that is divisors which are in the principal subgroup) $A_1/C_1,\cdots,A_n/C_n$ and the linear dependence of the functions $f_{A_1B_1C_1},\cdots,f_{A_nB_nC_n}$ over the field Φ, for all relevant choices of the B_i.

THEOREM 6. 1. *If the principal divisors $A_1/C_1,\cdots,A_n/C_n$ are dependent, then the functions $f_{A_1B_1C_1},\cdots,f_{A_nB_nC_n}$, for all relevant choices of the B_i, are linearly dependent over Φ.*

Proof. By induction over n.

1) If $n=2$, then since A_1/C_1 and A_2/C_2 are dependent, they are equal. Hence $A_1C_2=A_2C_1$. From Theorem 5. 2, combined with Theorem 5. 6,

$$f_{A_1B_1C_1}=f_{A_1C_2,B_1C_2,C_1C_2}=f_{A_2C_1,B_1C_2,C_2C_1}=\lambda f_{A_2C_1,B_2C_1,C_2C_1}=\lambda f_{A_2B_2C_2},\qquad \lambda\,\varepsilon\,\Phi.$$

Thus the theorem is true for $n=2$.

2) $n=3$.

Since $A_1/C_1,\cdots,A_3/C_3$ are dependent, $A_1C_2C_3$, $A_2C_1C_3$, $A_3C_1C_2$ are dependent. If any two of them are equal, the result will follow from the case $n=2$. Suppose the three are distinct. Without loss of generality each pair A_i, C_i are relatively prime integral divisors. Thus if $A_1C_2C_3$, $A_3C_1C_2$ and $C_1C_2C_3$ are dependent, then $C_1=C_3$; and then A_1, A_3 and C_1 are dependent. Thus by Theorem 5. 7, $f_{A_1B_1C_1}=\lambda f_{A_3B_3C_3}+\mu$ where $\lambda,\mu\,\varepsilon\,\Phi$. Their dependence would also force the dependence of $A_2C_1C_3$, $A_3C_1C_2$ and $C_1C_2C_3$. Thus $C_2=C_3$ and A_2, A_3, and C_1 are dependent. Thus $f_{A_2B_2C_2}=\alpha_{A_3B_3C_3}+\beta$ where $\alpha,\beta\,\varepsilon\,\Phi$. Combining the two expressions obtained, the linear dependence of the functions over Φ follows.

So let us assume that $A_1C_2C_3$, $A_3C_1C_3$, and $C_1C_2C_3$ are independent. Thus gram 4 all lie in the plane determined by these three g $B'_1C_2C_3$ and $B'_3C_1C_2$ as obtained in the diagram, 5. 6, that $f_{A_1B'_1C_1} = \lambda f_{A_1B_1C_1}$ and $f_{A_3B'_3C_3} = \mu f_{A_3B_3C_3}$ m the definition of the sum of two functions, $= \lambda f_{A_1B_1C_1} + \mu f_{A_3B_3C_3}$ and hence the theorem for

rem is true for all $n < r$ $(r > 3)$.

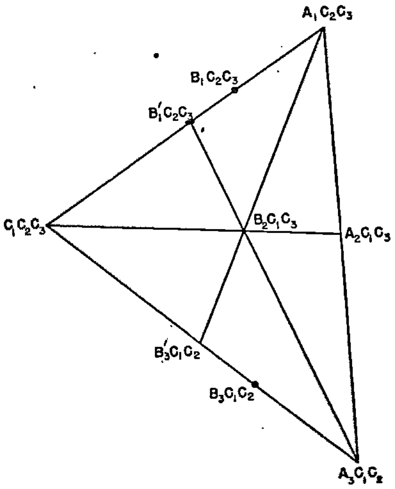


DIAGRAM 4.

Consider $A_1/C_1, \cdots, A_r/C_r$, a sequence of dependent principal divisors. If any $r-1$ of them or less are dependent, then we are done. Hence we suppose any $r-1$ of them or less are independent. Since $r > 3$, A_1/C_1, A_2/C_2 are independent and $A_3/C_3, \cdots, A_r/C_r$ are independent. But since the A_i/C_i, $i = 1, \cdots, r$ are dependent, by axiom 3 c. there is a divisor R/T which depends on both A_1/C_1, A_2/C_2 and on $A_3/C_3, \cdots, A_r/C_r$. Hence by the induction hypothesis f_{RST} is linearly dependent (over Φ) on $f_{A_1B_1C_1}$, $f_{A_2B_2C_2}$ and $f_{A_3B_3C_3}, \cdots, f_{A_rB_rC_r}$. Thus the $f_{A_iB_iC_i}$, $i = 1, \cdots, r$, are linearly dependent over Φ; completing the induction.

THEOREM 6. 2. *If* $f_{A_1B_1C_1}, \cdots, f_{A_nB_nC_n}$ *are linearly dependent over* Φ, *then* $A_1/C_1, \cdots, A_n/C_n$ *are dependent divisors.*

Proof. By induction over n:

1) If $n = 2$, $f_{A_1B_1C_1} = \lambda f_{A_2B_2C_2} = f_{A_2B'_2C_2}$ by Theorem 5. 6, and so by Theorem 5. 10, $A_1/C_1 = A_2/C_2$.

2) If $n = 3$ then by hypothesis $\sum_1^3 \lambda_i f_{A_iB_iC_i} =$

the result follows from the case $n = 2$. Let us assu
Then by Theorem 5.6 $\lambda_i f_{A_iB_iC_i} = f_{A_iB'_iC_i}$ and $-f_{A_1}$
$+ f_{A_3B'_3C_3}$. Constructing the sum $f_{A_2B'_2C_2} + f_{A_3B'_3C_3}$ we o
(not a constant since it must be equal to $f_{A_1B''_1C_1}$ wl
where, in all cases of addition, R/C_2C_3, A_2/C_2, A_3/C
$f_{A_1B''_1C_1} = f_{R,S,C_2C_3}$, by Theorem 5.10 $A_1/C_1 = R/C$
divisors A_1/C_1, A_2/C_2 and A_3/C_3 are dependent.

3) Suppose the theorem is true for $n = r - 1 \geq 3$.

Suppose that $\sum_1^r \lambda_i f_{A_iB_iC_i} = 0$, and not all the $\lambda_i = 0$. If any $\lambda_i = 0$ then the result follows from the induction hypothesis. So we may assume that no $\lambda_i = 0$. Let $f_{RST} = \sum_1^{r-2} \lambda_i f_{A_iB_iC_i}$. If f_{RST} is not a constant, then by the induction hypothesis, and since no $\lambda_i = 0$, R/T depends on $A_1/C_1, \cdots, A_{r-2}/C_{r-2}$. Since $f_{RST} + \lambda_{r-1} f_{A_{r-1}B_{r-1}C_{r-1}} + \lambda_r f_{A_rB_rC_r} = 0$, from the case $n = 3$, R/T depends on A_{r-1}/C_{r-1} and A_r/C_r. Thus A_r/C_r depends on A_{r-1}/C_{r-1} and R/T and so on $A_{r-1}/C_{r-1}, A_1/C_1, \cdots, A_{r-2}/C_{r-2}$. Thus in this case the theorem would be true. So suppose that f_{RST} is a constant. Thus, from Theorems 5.5 and 5.7 we may assume that $C_r = C_{r-1}$ and $A_{r-1} \varepsilon [A_r, C_r]$ since $f_{A_rB_rC_r} = \alpha f_{A_{r-1}B_{r-1}C_{r-1}} + \beta$ where $\alpha, \beta \varepsilon \Phi$. Now if $f_{R'S'T} = \sum_2^{r-1} \lambda_i f_{A_iB_iC_i}$ and if $f_{R'S'T'}$ is not a constant the result will follow as above. So suppose that $\lambda_1 f_{A_1B_1C_1} + \lambda_{rA_rB_rC_r} = \alpha$ in Φ. Hence again we may assume that $C_1 = C_r$ and $A_1 \varepsilon [A_r, C_r]$. Thus A_1, A_{r-1}, A_r are dependent$_1$ and since the three C's are equal, $A_1/C_1, A_{r-1}/C_{r-1}, A_r/C_r$ are dependent. Thus we have proved the theorem.

Combining Theorems 6.1 and 6.2 we have

THEOREM 6.3. *The functions $f_{A_1B_1C_1}, \cdots, f_{A_nB_nC_n}$ are linearly dependent over Φ if and only if the divisors $A_1/C_1, \cdots, A_n/C_n$ are dependent.*

7. Algebraic properties of K_G.

We will now consider some of the algebraic properties of K_G; in fact we will show that K_G is an algebraic function field in one variable over the algebraically closed field Φ. Moreover we show that if K is an algebraic function field in one variable over an algebraically closed constant field k, and $\mathfrak{G}$ is the divisor algebra of K, then K is isomorphic to K_G.

lowing lemma:

integral divisor then $-1 \leq n(A) - \dim\{A\}$.

3.4, if $A \mid B$ then $n(A) - \dim\{A\} \leq n(B)$
al then $I \mid A$ and so the result follows immediately.

$_C\ \varepsilon\ K_G$ *then* f_{ABC} *is transcendental over* Φ.

by the corollary to Theorem 3.6, for all $n > 0$
independent divisors. Hence by Theorem 6.3 the functions $f_{A^nB_nC^n}, \cdots, f_{ABC}$ are linearly independent over Φ, for all n. From the definition of multiplication $f_{A^iB_iC^i} = \lambda_1 f^i{}_{ABC}$. Thus we have shown that the positive integral powers of f_{ABC} are linearly independent over Φ.

THEOREM 7.2. *If A and B, and also C and D are independent integral divisors, then there exist integers m and n such that the divisors A^iC^j/B^iD^j, $i = 0, 1, \cdots, m$, $j = 0, 1, \cdots, n$ are dependent. In fact $n = n(A)$ and $m = (n(C) - 1)n(A) + 1$ will do.*

Proof. We multiply the sequence $\{A^i/B^i \cdot C^j/D^j\}$ by B^mD^m. We then obtain a sequence $\{A^iB^{m-i}C^jD^{n-j}\}$ of integral divisors. If for some $i \neq i_0$ or $j \neq j_0$ $A^iB^{m-i}C^jD^{n-j} = A^{i_0}B^{m-i_0}C^{j_0}D^{n-j_0}$, then $(A/B)^{i-i_0} = (C/D)^{j-j_0}$. Thus they would be dependent and the theorem would be true. If they are all distinct then we would have $mn + m + n + 1$ distinct integral divisors. If we could force $\dim\{A^mC^n\} < mn + m + n + 1$ for some m and n, then the divisors $A^iB^{m-i}C^jD^{n-j}$ would be dependent. By Lemma 4 $\dim\{A^mC^n\} \leq n(A^mC^n) + 1$. Thus if for some m and n, $n(A^mC^n) < mn + m + n$, we are finished. Hence we try to satisfy $m \cdot n(A) + n \cdot n(C) < mn + m + n$. We see that $n = n(A)$ and $m = (n(C) - 1)(n(A)) + 1$ will do.

THEOREM 7.3. *If $X, Y\ \varepsilon\ K_G$ then there exist a finite number of $a_{ij}\ \varepsilon\ \Phi$, not all 0, such that $\sum a_{ij}X^iY^j = 0$.*

Proof. If either X or Y is in Φ the theorem is trivial. Thus we may assume that $X = f_{ABC}$ and $Y = f_{DEF}$ where $A, B, \cdots, F$ are all integral divisors. From Theorem 7.2 there exist integers m and n such that the divisors $(A^i/C^i)(D^j/F^j)$ with $i = 0, 1, \cdots, m$ and $j = 0, 1, \cdots, n$, are dependent. Thus the functions $f_{A^iD^j,\ v_{ij},\ C^iF^j}$ are linearly dependent over Φ. Thus from the definition of the product of two functions, the functions $f^i{}_{ABC}f^j{}_{DEF}$ are linearly dependent over Φ.

COROLLARY 1. *K_G is of degree of transcendence one over Φ.*

Definition. An extension field K of a field k will be sai
and of bounded degree over k if there exists an integer N
every $x \varepsilon K$ there exists a polynomial p_x of degree at mos
in k and such that $p_x(x) = 0$.

COROLLARY 2. *K_G is algebraic and of bounded deg*

Proof. From Theorem 7.3 K_G is algebraic over $\Phi(f_A$
are relatively prime integral divisors, as we may assume, then
$N = n(A)$ will satisfy the requirements of the above definit

THEOREM 7.4. *Φ is algebraically closed.*

Proof. To show the algebraic closure it is sufficient to show the reducibility of all non-linear polynomials in $\Phi[x]$ for x any element transcendental over Φ. Pick $x = f_{ABC}$, and let $\sum_{i=0}^{n} a_i x^i$ be any polynomial in $\Phi[x]$. Construct the sum $\sum a_i f^i{}_{ABC} = f_{R,S,C^n}$; by Theorem 7.1 $R \neq C^n$. Without loss of generality R and C^n are relatively prime, for otherwise we could take out the common factor. Thus we can find a P such that $P \mid R$, $P \nmid C^n$. Thus $f_{R,S,C^n}(P) = 0$, and so f_{R,S,C^n} has as a linear factor, $f_{ABC} - f_{ABC}(P)$. Hence the theorem.

THEOREM 7.5.[4] *If K is transcendental over the perfect field k and K is algebraic and of bounded degree over $k(x)$, then K is a finite extension of $k(x)$.*

Proof. If k is of characteristic zero then K is separable over $k(x)$ and the result is immediate.

Suppose k is of characteristic $p \neq 0$. Then since k is perfect it contains the p-th roots of all its elements. Let S be the set of all f in K which are separable over $k(x)$. Then, by Van der Waerden [3, p. 129] S is a field. Since S is separable, algebraic and of bounded degree over $k(x)$, it is a finite, and hence simple, extension of $k(x)$. Let $S = k(x, y)$. Since K is algebraic and of bounded degree over $k(x)$ and k is perfect, for sufficiently large r, $S(x^t, y^t) \supset K$ where $t = 1/p^r$. But the degree of $S(x^t, y^t)$ over S is at most t^2 and hence, because S is a finite extension of $k(x)$, $S(x^t, y^t)$ is a finite extension of $k(x)$. Since $K \subset S(x^t, y^t)$, K is also finite over $k(x)$.

Since in our case K_G is algebraic and of bounded degree over $\Phi(f_{ABC})$ and Φ is algebraically closed, and hence perfect, we immediately have

---

[4] The author is indebted to Professor George Whaples for this simple proof of the theorem.

εK_G *then* K_G *is a finite extension of* $\Phi(f_{ABC})$.

ition [1, chapter 1] K_G is an algebraic function

e concrete case of an algebraic function field K
lgebraically closed field k. Suppose that 𝔖 is
etrized as in Section 1. Thus we know that k
YZ (Theorem 1.1) and we pick a definite one Φ.
Section 1) we define $\sigma^*(f) = f_{ABC}$. If σ is the isomorphism of k onto Φ, we define $\sigma^*(a) = \sigma(a)$ for a in k. Thus from Theorems 1.2 and 1.3 we have

THEOREM 7.7. *K is isomorphic to K_G under the isomorphism σ^*.*

Before closing we would like to point out that in the definitions of sums and products of our functions the concept of prime divisor was not used. Thus if we consider a projective space whose elements form a group and where group multiplication preserves collinearity, such sums and products (of marked lines) could be defined. The prime divisors were used to justify that these definitions of sum and product lead to a field.

INDIANA UNIVERSITY.

BIBLIOGRAPHY.

[1] S. MacLane, Lecture notes on "*Algebraic Functions*," Cambridge 1947.
[2] Hodge and Pedoe, "*Methods of Algebraic Geometry*," vol. 1, Cambridge 1947.
[3] B. L. Van der Waerden, "*Moderne Algebra*," vol. 1, Second Edition, Berlin 1940.

PRIME IDEALS IN GENERAL RINGS.

By NEAL H. McCOY.

1. Introduction. The concept of prime ideal has pla
role in the theory of commutative rings, but has not been u
in the study of noncommutative rings. Some properties o
general rings have been discussed by Krull [9] and by Fitting [3]. However, except in these papers, prime ideals seem to have been used only incidentally and not made the subject of special study. It is the purpose of the present paper to extend to general, that is, not necessarily commutative, rings several results which are well known in the commutative case.

Unless otherwise stated, the word *ideal* shall mean *two-sided* ideal. In a *commutative* ring R an ideal $\mathfrak{p}$ is a *prime* ideal if and only if $ab \equiv 0(\mathfrak{p})$ implies that $a \equiv 0(\mathfrak{p})$ or $b \equiv 0(\mathfrak{p})$. Naturally, this definition could also be used in noncommutative rings, as has been pointed out by Fitting [3], who says that a prime ideal according to this definition is *completely* prime. However, it turns out that this concept is not particularly useful, since a noncommutative ring seldom contains very many completely prime ideals. In other words, the defining condition is too strong to be of much interest.

In an arbitrary ring R, it is customary to call an ideal $\mathfrak{p}$ a *prime ideal* if and only if $\mathfrak{a}\mathfrak{b} \equiv 0(\mathfrak{p})$ implies that $\mathfrak{a} \equiv 0(\mathfrak{p})$ or $\mathfrak{b} \equiv 0(\mathfrak{p})$, it being understood that $\mathfrak{a}$ and $\mathfrak{b}$ are ideals in R. An ideal which is completely prime is prime, but the converse is not generally true. However, these concepts coincide in the case of commutative rings.

Our first theorem gives a number of properties of an ideal, each of which is equivalent to that just used to define a prime ideal. In particular, it follows that an ideal $\mathfrak{p}$ in the arbitrary ring R is a prime ideal in R if and only if $aRb \equiv 0(\mathfrak{p})$ implies that $a \equiv 0(\mathfrak{p})$ or $b \equiv 0(\mathfrak{p})$. This suggests the desirability of defining an *m-system* (generalizing the familiar concept of multiplicative system) M of elements of R as a system with the property that $c \varepsilon M$, $d \varepsilon M$ imply the existence of an element x of R such that $cxd \varepsilon M$. Thus an ideal $\mathfrak{p}$ in R is a prime ideal if and only if the complement of $\mathfrak{p}$ in R is an m-system. This characterization of the prime ideals plays an important role in the sequel.

---

* Received July 20, 1948; presented to the American Mathematical Society, April 17, 1948.

R, the *radical* of the ideal $\mathfrak{a}$ is defined to be the set
ith the property that every m-system which contains r
$\mathfrak{a}$.[1] It is shown in **3** that the radical of $\mathfrak{a}$ is the inter-
e ideals which contain $\mathfrak{a}$. The methods are based on
l the results reduce to those of Krull if R happens to be
he material of **3** is a simple adaptation of the exposition
found in Chapter V of [14].

nber of different definitions of the radical of a general ring have been proposed. We shall add to this list by giving still another definition as follows. The *radical* N of the ring R is the radical of the zero ideal in R. We shall show that N is a nil ideal which contains every nilpotent ideal of R, and that N is a *radical ideal* in the sense of Baer [1]. The relation of N to the radicals of Kothe [8] and Levitzki [11] and [12] is still an unsolved problem. In common with all the other definitions of the radical of a general ring, N becomes the classical radical in the presence of the descending chain condition for right ideals. Furthermore, N has all the usual properties expected of a radical.

A *primitive* ideal as defined by Jacobson [7] is a prime ideal, and hence N is contained in the Jacobson radical of R. We may also point out that the method used by Jacobson [7] to introduce a topology in the set of primitive ideals in a ring can be used without modification to introduce a topology in the set of prime ideals in a ring. In fact, several of the results of [7] can be easily carried over to results about the space of prime ideals in a ring.

The radical recently defined by Brown and McCoy [2] is also the intersection of a certain class of prime ideals, namely, those maximal ideals $\mathfrak{m}$ such that $R/\mathfrak{m}$ has a unit element.

A ring in which (0) is a prime ideal may be called a *prime ring*. Thus the primitive rings of Jacobson [6] are prime rings. In **5** we shall prove that a prime ring which contains minimal right ideals is a primitive ring. However, these concepts do not coincide in general, for any integral domain is a prime ring and an integral domain is primitive if and only if it is a field.

We shall point out in Theorem 6 that a ring is isomorphic to a subdirect sum of prime rings if and only if it has zero radical. This is an

---

[1] Fitting [3] defined the radical of $\mathfrak{a}$ to be the set of elements which generate nil ideals modulo $\mathfrak{a}$. The radical of $\mathfrak{a}$ as defined above is contained in Fitting's radical, but the exact relation between these concepts is an unsolved problem.

analogue of one of the Wedderburn-Artin structure
this result it would seem desirable to make a further st

2. Definition and fundamental properties. W
following result:

THEOREM 1. *If* $\mathfrak{p}$ *is an ideal in the arbitrary*
conditions are equivalent:

(i) *If* $\mathfrak{a}$, $\mathfrak{b}$ *are ideals in* R *such that* $\mathfrak{a}\mathfrak{b} \equiv 0(\mathfrak{p})$, *then* $\mathfrak{a} \equiv 0(\mathfrak{p})$, *or* $\mathfrak{b} \equiv 0(\mathfrak{p})$.

(ii) *If* (a), (b) *are principal ideals in* R *such that* $(a)(b) \equiv 0(\mathfrak{p})$, *then* $a \equiv 0(\mathfrak{p})$ *or* $b \equiv 0(\mathfrak{p})$.

(iii) *If* $aRb \equiv 0(\mathfrak{p})$, *then* $a \equiv 0(\mathfrak{p})$ *or* $b \equiv 0(\mathfrak{p})$.

(iv) *If* I_1, I_2 *are right ideals in* R *such that* $I_1 I_2 \equiv 0(\mathfrak{p})$, *then* $I_1 \equiv 0(\mathfrak{p})$ *or* $I_2 \equiv 0(\mathfrak{p})$.

(v) *If* J_1, J_2 *are left ideals in* R *such that* $J_1 J_2 \equiv 0(\mathfrak{p})$, *then* $J_1 \equiv 0(\mathfrak{p})$ *or* $J_2 \equiv 0(\mathfrak{p})$.

Before giving the proof we make one observation which will be useful. Clearly (i), although stated for the product of two ideals, implies that if a product of any finite number of ideals is in $\mathfrak{p}$, at least one of the ideals is in $\mathfrak{p}$. A similar result holds for (ii), but is not quite so obvious. However, suppose that (ii) holds and that $(a)(b)(c) \equiv 0(\mathfrak{p})$ with $a \not\equiv 0(\mathfrak{p})$. Then for *every* b_1 in (b), c_1 in (c), we have $(a)(b_1 c_1) \equiv 0(\mathfrak{p})$, which then implies that $b_1 c_1 \equiv 0(\mathfrak{p})$. This shows that $(b)(c) \equiv 0(\mathfrak{p})$, and hence $b \equiv 0(\mathfrak{p})$ or $c \equiv 0(\mathfrak{p})$. In like manner, the result can be established for the product of any finite number of principal ideals.

We are now ready to prove the theorem. Clearly (i) implies (ii). We now assume (ii) and prove (iii). Suppose that $aRb \equiv 0(\mathfrak{p})$, from which it follows that $RaRbR \equiv 0(\mathfrak{p})$, and thus $(a)^2(b)^3 \subseteq RaRbR \equiv 0(\mathfrak{p})$. By the observation made above, (ii) implies that $a \equiv 0(\mathfrak{p})$ or $b \equiv 0(\mathfrak{p})$, and this establishes (iii).

Now let us assume (iii) and suppose that I_1, I_2 are right ideals such that $I_1 I_2 \equiv 0(\mathfrak{p})$ with $I_1 \not\equiv 0(\mathfrak{p})$. Let a_1 be an element of I_1 not in $\mathfrak{p}$. Then for *every* element a_2 of I_2 we have $a_1 R a_2 \subseteq I_1 I_2 \equiv 0(\mathfrak{p})$. Hence, by (iii), hus $I_2 \equiv 0(\mathfrak{p})$, and we have therefore shown that milar argument will show that also (iii) implies (v). ted by observing that (i) is implied by either (iv)

ideal $\mathfrak{p}$ with any one (and therefore all) of the orem 1 is a *prime ideal.*

a prime ideal in R, and a an element of R such that $RaR \equiv 0(\mathfrak{p})$, *then* $a \equiv 0(\mathfrak{p})$.

To prove this, we observe that $RaR \equiv 0(\mathfrak{p})$ implies that $aRaR \equiv 0(\mathfrak{p})$, and (iv) shows that $aR \equiv 0(\mathfrak{p})$. It then follows that $aRa \equiv 0(\mathfrak{p})$, and we must have $a \equiv 0(\mathfrak{p})$ by (iii).

We next prove the following result:

LEMMA 2. *If $\mathfrak{b}$ is an ideal in R, and $\mathfrak{p}$ a prime ideal in R, then $\mathfrak{b} \cap \mathfrak{p}$ is a prime ideal in the ring $\mathfrak{b}$.*

Let b_1, b_2 be elements of $\mathfrak{b}$ such that $b_1\mathfrak{b}b_2 \equiv 0(\mathfrak{p} \cap \mathfrak{b})$. Then $b_1Rb_2Rb_2 \subseteq b_1\mathfrak{b}b_2 \equiv 0(\mathfrak{p})$, and hence $b_1Rb_2Rb_2R \equiv 0(\mathfrak{p})$. From this, (iv) implies that $b_1R \equiv 0(\mathfrak{p})$ or $b_2R \equiv 0(\mathfrak{p})$. If $b_1R \equiv 0(\mathfrak{p})$, then $b_1Rb_1 \equiv 0(\mathfrak{p})$ and (iii) implies that $b_1 \equiv 0(\mathfrak{p})$. Similarly, if $b_2R \equiv 0(\mathfrak{p})$, we have $b_2 \equiv 0(\mathfrak{p})$. Thus either $b_1 \equiv 0(\mathfrak{p} \cap \mathfrak{b})$ or $b_2 \equiv 0(\mathfrak{p} \cap \mathfrak{b})$, and $\mathfrak{p} \cap \mathfrak{b}$ is a prime ideal in the ring $\mathfrak{b}$ by (iii).

Definition 2. A set M of elements of R is an *m-system* if and only if $c \in M$, $d \in M$ imply that there exists an element x of R that $cxd \in M$. The void set is to be considered as an m-system.

The importance of this concept lies in the fact that, by (iii), an ideal $\mathfrak{p}$ in R is a prime ideal if and only if its complement $C(\mathfrak{p})$ in R is an m-system. The agreement to consider the void set as an m-system is to take care of the special case in which $\mathfrak{p} = R$, for clearly R is a prime ideal in R.

It will be observed that the concept of an m-system is a generalization of that of multiplicative system. For if M is a multiplicative system with $c \in M$, $d \in M$, then there is an element x (either c or d may be used) such that $cxd \in M$, and hence M is an m-system according to the definition given above.

3. The radical of an ideal. This section is base
of Chapter V of [14] which, in turn, is largely an exp
to Krull [10].

Definition 3. The *radical* $\mathfrak{r}$ of an ideal $\mathfrak{a}$ in R cons
r of R with the property that every m-system which c
element of $\mathfrak{a}$.

It will presently appear that $\mathfrak{r}$ is an ideal in R. However, we first observe that $\mathfrak{a} \subseteq \mathfrak{r}$. Furthermore, $\mathfrak{a}$ and $\mathfrak{r}$ are contained in precisely the same prime ideals. For suppose that $\mathfrak{a} \subseteq \mathfrak{p}$, where $\mathfrak{p}$ is a prime ideal, and that $r \varepsilon \mathfrak{r}$. If r were not in $\mathfrak{p}$, that is, if $r \varepsilon C(\mathfrak{p})$, then $C(\mathfrak{p})$ would have to contain an element of $\mathfrak{a}$ since $C(\mathfrak{p})$ is an m-system. But clearly $C(\mathfrak{p})$ contains no element of $\mathfrak{a}$, and therefore r is not in $C(\mathfrak{p})$. Thus $r \varepsilon \mathfrak{p}$, and hence $\mathfrak{r} \subseteq \mathfrak{p}$ as required.

Definition 4. A prime ideal $\mathfrak{p}$ is a *minimal prime ideal belonging to the ideal* $\mathfrak{a}$ if and only if $\mathfrak{a} \subseteq \mathfrak{p}$ and there exists no prime ideal $\mathfrak{p}'$ such that $\mathfrak{a} \subseteq \mathfrak{p}' \subset \mathfrak{p}$.[2]

We are now ready to state the principal theorem of this section as follows:

THEOREM 2. *The radical* $\mathfrak{r}$ *of an ideal* $\mathfrak{a}$ *is the intersection of all the minimal prime ideals belonging to* $\mathfrak{a}$.

We shall establish several lemmas and then show how they lead to an immediate proof of the theorem.

If two sets of elements of R have no elements in common, we may say that either of these sets *does not meet* the other.

LEMMA 3. *Let* $\mathfrak{a}$ *be an ideal in* R, *and* M *an m-system which does not meet* $\mathfrak{a}$. *Then* M *is contained in an m-system* M' *which is maximal in the class of m-systems which do not meet* $\mathfrak{a}$.

This is, of course, an immediate consequence of Zorn's Maximum Principle and is merely stated in the form of a lemma for convenience of reference.

LEMMA 4. *Let* M *be an m-system in* R, *and* $\mathfrak{a}$ *an ideal which does not*

---

[2] By $\mathfrak{p}' \subset \mathfrak{p}$ we mean that $\mathfrak{p}'$ is properly contained in $\mathfrak{p}$.

in an ideal $\mathfrak{p}^*$ which is maximal in the class
I. The ideal $\mathfrak{p}^*$ is necessarily a prime ideal.

rs at once from the Maximum Principle. We
leal. Suppose that $a \not\equiv 0(\mathfrak{p}^*)$ and $b \not\equiv 0(\mathfrak{p}^*)$.
f $\mathfrak{p}^*$ implies that $(\mathfrak{p}^*, a)$ contains an element
) contains an element m_2 of M. Thus there
of (b) such that $m_1 \equiv a_1(\mathfrak{p}^*)$, $m_2 \equiv b_1(\mathfrak{p}^*)$.
is an element x of R such that $m_1 x m_2 \, \varepsilon \, M$, and
* does not meet M. But $a_1 x b_1 \equiv m_1 x m_2(\mathfrak{p}^*)$
and therefore $a_1 x b_1 \not\equiv 0(\mathfrak{p}^*)$. However, $(a)(b)$ contains the element $a_1 x b_1$, and thus $(a)(b) \not\equiv 0(\mathfrak{p}^*)$. By property (ii) of Theorem 1, this shows that $\mathfrak{p}^*$ is a prime ideal.

We now prove

LEMMA 5. *A set $\mathfrak{p}$ of elements of the ring R is a minimal prime ideal belonging to $\mathfrak{a}$ if and only if $C(\mathfrak{p})$ is maximal in the class of m-systems which do not meet $\mathfrak{a}$.*

First, let $\mathfrak{p}$ be a set of elements of R with the property that $M = C(\mathfrak{p})$ is a maximal m-system which does not meet $\mathfrak{a}$. If $\mathfrak{p}^*$ is the prime ideal whose existence is asserted in Lemma 4, then $C(\mathfrak{p}^*)$ is an m-system which contains M and does not meet $\mathfrak{a}$. The maximal property of M implies that $C(\mathfrak{p}^*) = M = C(\mathfrak{p})$, and hence $\mathfrak{p} = \mathfrak{p}^*$. Thus $\mathfrak{p}$ is a prime ideal containing $\mathfrak{a}$. Clearly, there can exist no prime ideal $\mathfrak{p}_1$ such that $\mathfrak{a} \subseteq \mathfrak{p}_1 \subset \mathfrak{p}$, since this would imply that $C(\mathfrak{p}_1)$ is an m-system which does not meet $\mathfrak{a}$ and properly contains M. This is impossible because of the maximal property of M; hence $\mathfrak{p}$ is a minimal prime ideal belonging to $\mathfrak{a}$.

Conversely, if $\mathfrak{p}$ is a minimal prime ideal belonging to $\mathfrak{a}$, $M = C(\mathfrak{p})$ is an m-system which does not meet $\mathfrak{a}$, and Lemma 3 shows the existence of a maximal m-system M' which contains M and does not meet $\mathfrak{a}$. By the part of the theorem just proved, $C(M') = \mathfrak{p}'$ is a minimal prime ideal belonging to $\mathfrak{a}$. Since $M' \supseteq M$, it follows that $\mathfrak{p}' \subseteq \mathfrak{p}$. Thus $\mathfrak{a} \subseteq \mathfrak{p}' \subseteq \mathfrak{p}$, from which it follows that $\mathfrak{p} = \mathfrak{p}'$, and thus $M = M'$. This shows that $C(\mathfrak{p}) = M$ is a maximal m-system which does not meet $\mathfrak{a}$, and completes the proof of the lemma.

We are now ready to prove the theorem. If $\mathfrak{r}$ is the radical of $\mathfrak{a}$, we

have pointed out above that $\mathfrak{r}$ is contained in the same
This shows that $\mathfrak{r}$ is contained in the intersection of all
ideals belonging to $\mathfrak{a}$. Now let a be an element of R n
the definition of $\mathfrak{r}$, there exists an m-system M which
not meet $\mathfrak{a}$. By Lemma 3, M is contained in a maxima
does not meet $\mathfrak{a}$. By Lemma 5, $C(M')$ is a minimal p
to $\mathfrak{a}$, and clearly $C(M')$ does not contain a. Hence a can
section of all the minimal prime ideals belonging to $\mathfrak{a}$,
therefore established.

The following result is an immediate consequence of the theorem just proved:

COROLLARY. *The radical of an ideal is an ideal.*

If $\mathfrak{p}$ is any prime ideal containing $\mathfrak{a}$, then $M = C(\mathfrak{p})$ is an m-system which does not meet $\mathfrak{a}$. If M' is the m-system defined in Lemma 3, Lemma 5 shows that $C(M')$ is a minimal prime ideal belonging to $\mathfrak{a}$. Since $C(\mathfrak{p}) \subseteq M'$, it follows that $\mathfrak{a} \subseteq C(M') \subseteq \mathfrak{p}$. This proves that *any prime ideal which contains* $\mathfrak{a}$ *contains a minimal prime ideal belonging to* $\mathfrak{a}$.

4. The radical of a ring.

We now make the following definition:

Definition 5. The *radical of the ring* R is the radical of the zero ideal in R.

We shall henceforth denote the radical of the ring R by N. It is clear that N is a nil ideal, for if $a \varepsilon N$; the m-system $\{a, a^2, a^3, \cdots\}$ must contain 0, and a is therefore nilpotent. Furthermore, every element b which generates a nilpotent ideal (right, left, or two-sided) is in N. For if I is an ideal such that $I^n = 0$, then $I^n \equiv 0(\mathfrak{p})$ for every prime ideal $\mathfrak{p}$ in R, and this implies that $I \equiv 0(\mathfrak{p})$. Hence $I \equiv 0(N)$, since N is the intersection of all the prime ideals in R.

If $a \varepsilon N$, then clearly $aR \subseteq N$. Conversely, if $aR \subseteq N$, $RaR \subseteq N$, and Lemma 1 shows that $a \varepsilon N$. We see therefore *that* $aR \subseteq N$ *if and only if* $a \varepsilon N$.

THEOREM 3. *In the presence of the descending chain condition for right ideals, N coincides with the classical radical of R.*

If $a \varepsilon N$, (a) is a nil ideal, and it is known that the descending chain condition implies that (a) is then a nilpotent ideal. On the other hand, it was pointed out above that N contains all elements which generate nil-

sts precisely of the elements which generate
r, is one of the familiar characterizations of
oof is completed.

ieal in R, the radical of the ring $\mathfrak{b}$ is $\mathfrak{b} \cap N$.

the *ring* $\mathfrak{b}$, Lemma 2 shows that $N' \subseteq \mathfrak{b} \cap N$.
very m-system in R which contains b contains
0. Thus, in particular, every m-system in $\mathfrak{b}$ which contains b contains 0. This means that $b \in N'$, and thus $\mathfrak{b} \cap N \subseteq N'$, completing the proof.

THEOREM 5. *If N is the radical of R, then R/N has zero radical.*

To prove this, let $\bar{a}$ be an element of the radical of R/N, and thus $\bar{a}$ is contained in all prime ideals in R/N. If $\bar{a} \neq 0$, $a \not\equiv 0(N)$, and hence a is not contained in some prime ideal $\mathfrak{p}$ in R. Since $\mathfrak{p} \supseteq N$, we have $R/\mathfrak{p} \cong (R/N)/(\mathfrak{p}/N)$, from which it follows that $\mathfrak{p}/N$ is a prime ideal in R/N. Furthermore, $\mathfrak{p}/N$ does not contain $\bar{a}$ since $a \not\equiv 0(\mathfrak{p})$. This contradiction shows that we must have $\bar{a} = 0$, which completes the proof of the theorem.

It follows from this theorem that R/N contains no nonzero nilpotent ideals (right, left, or two-sided), for every nilpotent ideal in R/N must be in the radical of R/N. In particular, this shows that N is a *radical ideal* in the sense of Baer [1].

5. **Prime rings.** We shall now make the following

Definition 6. A ring R is a *prime ring* if and only if (0) is a prime ideal in R.

Theorem 1 yields a number of equivalent characterizations of the prime rings, one of the most interesting being that a ring R is a prime ring if and only if $aRb = 0$ implies that $a = 0$ or $b = 0$.

It is easy to see that a *commutative prime ring* is just an integral domain. Any simple ring S (with $S^2 \neq 0$) is a prime ring, and a primitive ring is also prime, as was shown by Jacobson [6]. From Lemma 2 we also observe that *an ideal in a prime ring is a prime ring.*

Now a prime ring has zero radical and hence in the presence of the descending chain condition for right ideals is isomorphic to a direct sum of

a finite number of simple rings. However, the (
simple rings is certainly not prime, and hence if
dition holds for right ideals, the concepts of pr
(with nonzero square) coincide.

If $\mathfrak{p}$ is a prime ideal in the arbitrary ring R
conversely. Since N is the intersection of all the p
argument[3] yields the following analogue of one
theorems:

THEOREM 6. *A necessary and sufficient condition that a ring be isomorphic to a subdirect sum of prime rings is that it have zero radical.*

This theorem indicates the importance of prime rings in the general structure theory. We shall now prove a few other results about prime rings.

THEOREM 7. *A prime ring that contains minimal right ideals is a primitive ring.*

The following simple proof is due to Bailey Brown. If I is a minimal right ideal of the prime ring R, then I is a simple R-module whose annihilator I^* in R is a right (in fact, two-sided) ideal such that $II^* = 0$. Since R is prime this implies that $I^* = 0$ and thus I is a simple R-module with zero annihilator, that is, R is isomorphic to an irreducible ring of endomorphisms. This implies[4] that R is primitive, and the proof is completed.

Now let T be a ring *with unit element,* and denote by T_n the ring of all matrices of order n with elements in T. We shall prove

THEOREM 8. *If T is a ring with unit element, then T_n is a prime ring if and only if T is a prime ring.*

As usual, let e_{ij} denote the matrix with the unit element in the i-th row and j-th column, and zeros elsewhere. If T is not prime, then T_n is not prime. For if T is not a prime ring, there exist nonzero elements a, b of T such that $aTb = 0$. This clearly implies that $(ae_{11})T_n(be_{11}) = 0$ with ae_{11} and be_{11} nonzero elements of T_n, and this shows that T_n is not a prime ring.

Conversely, suppose that T_n is not a prime ring, and hence that there exist nonzero matrices (a_{ij}), (b_{ij}) in T_n such that $(a_{ij})T_n(b_{ij}) = 0$. Let us assume that $a_{pq} \neq 0$, $b_{rs} \neq 0$. Now, for every x in T we must have

---

[3] See § 3 of [13] for references.

[4] Jacobson [6], p. 312.

$$[e_{qr}x][\sum_{k,l} b_{kl}e_{kl}] = \sum_{i,l} a_{iq}xb_{rl}e_{il} = 0.$$

ent of e_{ps} must be zero, that is, $a_{pq}xb_{rs} = 0$. Since
n T, this means that $a_{pq}Tb_{rs} = 0$, and T is not a
s therefore completed.
t we shall prove the following theorem about the

THEOREM 9. *If N is the radical of the arbitrary ring R, the radical of the complete matrix ring R_n is N_n.*

We first give the proof under the assumption that R has a unit element, and then remove this restriction. Since R is assumed to have a unit element, there is a one-to-one correspondence $M \leftrightarrow M_n$ between ideals in R and ideals in R_n. Furthermore, it is easily verified that $(R/M)_n \cong R_n/M_n$ and thus, by Theorem 8, M_n is a prime ideal in R_n if and only if M is a prime ideal in R. Thus if N is the radical of R, and $\mathfrak{p}_i$ are the prime ideals in R, we see that

$$\text{radical of } R_n = \bigcap (\mathfrak{p}_i)_n = (\bigcap \mathfrak{p}_i)_n = N_n.$$

If R does not have a unit element, it is well known that we can imbed R in a ring S with unit element in such a way that R is an ideal in S. If the radical of R is N, and the radical of S is N', then Theorem 4 shows that $N = R \cap N'$. By the result just proved, the radical of S_n is N'_n and, since R_n is an ideal in S_n, Theorem 4 shows that

$$\text{radical of } R_n = N'_n \cap R_n = (N' \cap R)_n = N_n,$$

thus completing the proof.

SMITH COLLEGE.

REFERENCES.

1. R. Baer, "Radical ideals," *American Journal of Mathematics*, vol. 65 (1943), pp. 537-568.

2. B. Brown and N. H. McCoy, "Radicals and subdirect sums," *American Journal of Mathematics*, vol. 69 (1947), pp. 46-58.

3. H. Fitting, " Primärkomponentenzerlegung in nichtkomn
matische Annalen, vol. 111 (1935), pp. 19-41.

4. O. Goldman, " Addition to my note on semi-simple
American Mathematical Society, vol. 53 (1947), p. 956.

5. N. Jacobson, " Structure theory of simple rings without
Transactions of the American Mathematical Society, vol. 57 (1

6. ———, " The radical and semi-simplicity for arbitrary r
of Mathematics, vol. 67 (1945), pp. 300-320.

7. ———, " A topology for the set of primitive ideals in an
ceedings of the National Academy of Sciences, vol. 31 (1945), pp.

8. G. Köthe, " Die Struktur der Ringe, deren Restklassenring nach dem Radikal vollständig reduzibel ist," *Mathematische Zeitschrift*, vol. 32 (1930), pp. 161-186.

9. W. Krull, " Zur Theorie der zweiseitigen Ideale in nichtkommutativen Bereichen," *Mathematische Zeitschrift*, vol. 28 (1928), pp. 481-503.

10. ———, " Idealtheorie in Ringen ohne Endlichkeitsbedingung," *Mathematische Annalen*, vol. 101 (1929), pp. 729-744.

11. J. Levitzki, " On the radical of a general ring," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 462-466.

12. ———, " On three problems concerning nil-rings," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 913-919.

13. N. H. McCoy, " Subdirect sums of rings," *Bulletin of the American Mathematical Society*, vol. 53 (1947), pp. 856-877.

14. ———, *Rings and Ideals*, Carus Mathematical Monographs, no. 8, 1948.

WITHOUT NILPOTENT IDEALS.*

By A. H. Clifford.

esent paper is to extend the results of a previous
aving a zero element. It was shown there that
at least one left and at least one right minimal
ideal, then it contains a unique minimal two-sided ideal $\mathfrak{R}$ (the "Suschkewitsch kernel," which we formerly denoted by K), and that $\mathfrak{R}$ is a completely simple semigroup without zero, in the sense of Rees [2]. If S has a zero element 0, then $\mathfrak{R} = (0)$, and this result becomes trivial. We alter our definition of minimal ideal to mean minimal but $\neq (0)$, and prove (Theorem 3.1 below) that if M is a minimal two-sided ideal of S containing minimal left and right ideals of S, and if S contains no nilpotent ideal $\neq (0)$, then M is a completely simple semigroup (with zero) [3].

A simple but basic lemma in ring theory is that if L is a minimal left ideal, then either $L^2 = (0)$ or L has an idempotent generator: $L = Se$ $(e^2 = e)$. This is not so for semigroups. An example is given by Baer and Levi [4] of a semigroup S which is left simple, hence is itself a minimal left ideal of S, but contains no idempotent element. But it is true for semigroups S having no nilpotent ideals $\neq (0)$, and in which every two-sided ideal contains minimal left and right ideals (Theorem 4.1).

In the concluding section we discuss the connection between these results and S. Schwarz's theory [5] of semigroups having a kernel $\mathfrak{R}$ and radical $\mathfrak{R}$. Schwarz calls an ideal A of S "$\mathfrak{R}$-potent" if some power A^n of A is contained in $\mathfrak{R}$, and defines the radical $\mathfrak{R}$ of S to be the sum of all the $\mathfrak{R}$-potent two-sided ideals of S. If $\mathfrak{R}$ is itself $\mathfrak{R}$-potent, then the difference semigroup $S - \mathfrak{R}$ as defined by Rees (*loc. cit.*, p. 389) has no non-zero nilpotent ideals, and hence the foregoing results may be applied thereto (Theorems 5.2 and 5.3). These are compared with similar results found by Schwarz. Finally it is remarked that, under the assumptions made, $\mathfrak{R}$ contains all nil-ideals of S.

1. Simplicity of minimal two-sided ideals. As customary in the calculus of complexes, the product $A_1 A_2 \cdots A_n$ of a finite number of subsets A_i of a semigroup S shall mean the set of all products $a_1 a_2 \cdots a_n$ with a_i in A_i $(i = 1, 2, \cdots, n)$. In particular, A^n shall mean the set of all products $a_1 a_2 \cdots a_n$ of n elements a_i of A.

---

* Received October 22, 1948.

A left [right] ideal of S is a non-vacuous subset A
$SA \subseteq A$ $[AS \subseteq A]$. A set A which is both a left and
a two-sided ideal. The class sum of any number of
ideals is a left [right, two-sided] ideal, and the sa
intersection.

Assume now that S contains a zero, i. e. an elem
$= 0 \cdot s = 0$ for all s in S. Clearly S can contain on
The single-element set (0) is a two-sided ideal of S contained or
right ideal of S. A left or right ideal A of S will be called nilpotent if $A^n = (0)$ for some positive integer n. We shall say that S is *without nilpotent ideals* if S contains a zero element but no nilpotent left or right ideal $\neq (0)$. (From Lemma 5.2 below, this condition is satisfied if S contains a zero element but no nilpotent two-sided ideal $\neq (0)$.)

A left [right, two-sided] ideal of S will be called a minimal left [right, two-sided] ideal of S if it is $\neq (0)$ and contains no proper subset $\neq (0)$ which is also a left [right, two-sided] ideal of S. If A and B are two distinct minimal left [right, two-sided] ideals of S, their intersection $A \cap B = (0)$. If A is a minimal two-sided ideal and B a minimal left or right ideal, either $A \cap B = (0)$ or $B \subseteq A$.

THEOREM 1.1. *Let S be a semigroup without nilpotent ideals. Then any minimal two-sided ideal of S is a simple semigroup.*

Proof. Let M be a minimal two-sided ideal of S, and suppose (by way of contradiction) that B is a proper ideal $\neq (0)$ of M. MBM is a two-sided ideal of S contained in B, hence properly contained in M, hence $= (0)$ by the minimality of M. Then $(MB)^2 = MBMB = (0)$, so that the left ideal MB of S is nilpotent. By hypothesis this requires $MB = (0)$. Similarly, $(BM)^2 = (0)$ and hence $BM = (0)$.

Now SBS is a two-sided ideal of S contained in M, so that either $SBS = (0)$ or M. In either event, since $MB = (0)$, we conclude that $(SB)^2 = SBS \cdot B = (0)$. Thus the left ideal SB of S is nilpotent, and hence $SB = (0)$. Similarly, $BS = (0)$. But then $SB \subseteq B$, $BS \subseteq B$, so that B is a two-sided ideal of S, contrary to the minimality of M.

2. Two-sided ideals containing minimal left ideals.

LEMMA 2.1. *Let S be a semigroup with zero 0. If L is a minimal left ideal of S, and c is any element of S, then either Lc is also a minimal left ideal of S or $Lc = (0)$.*

Proof. Lc is clearly a left ideal of S. Assume $Lc \neq (0)$, and suppose that B is a left ideal $\neq (0)$ of S contained in Lc. Let L_1 be the set of all

εB. $L_1 \neq (0)$ since otherwise $B = (0)$. For
$a_1 \varepsilon L_1$. Hence L_1 is a left ideal $\neq (0)$ of S
ideal L. This requires $L_1 = L$, $B \supseteq L_1 c = Lc$,

e a semigroup with zero 0. *Let A be a two-*
least one minimal left ideal of S. Then the
left ideals of S contained in A is a two-sided
ideal … *A is minimal, it is a sum of minimal left ideals of S.*

Proof. As a sum of left ideals, B is a left ideal. To show that it is also a right ideal, we must show that if $b \,\varepsilon\, B$ and $c \,\varepsilon\, S$, then $bc \,\varepsilon\, B$. By definition of B, b must belong to some minimal left ideal L of S contained in A. $bc \,\varepsilon\, Lc$, and, by Lemma 2.1, Lc is a minimal left ideal, evidently contained in A, or else $Lc = (0)$. In either event $Lc \subseteq B$, and hence $bc \,\varepsilon\, B$. Clearly, if A is minimal, then $A = B$.

THEOREM 2.2. *Let S be a semigroup without nilpotent ideals. Let M be a minimal two-sided ideal of S containing at least one minimal left ideal of S. Then every left ideal of M is a left ideal of S.*

Proof. We show first that any minimal left ideal L of S contained in M is also minimal regarded as a left ideal of M. Suppose that B is a left ideal $\neq (0)$ of M contained in L. MB is a left ideal of S contained in $ML \subseteq L$, whence $MB = L$ or $MB = (0)$. But $MB \subseteq B$, so that the first alternative would imply $L \subseteq B$, $B = L$, as desired. We may therefore assume $MB = (0)$. Now SB is a left ideal of S contained in $SL \subseteq L$, whence $SB = L$, or $SB = (0)$. The second alternative implies that B is itself a left ideal of S, whence $B = L$ from the minimality of L. We may therefore assume $SB = L$. Then $L^2 = SBSB \subseteq SMSB \subseteq MB = (0)$, contrary to the assumption that S contains no nilpotent left ideal $\neq (0)$.

Now let A be any left ideal of M. If $A = (0)$, A is clearly a left ideal of S, so we may assume $A \neq (0)$. By Theorem 2.1, M is the sum of all the minimal left ideals of S contained in it. Hence each element $a \neq 0$ of A belongs to some minimal left ideal L of S contained in M. $L \cap A$ is a left ideal $\neq (0)$ of M contained in L. Since, as we have just proved, L is also minimal regarded as a left ideal of M, we conclude $L \cap A = L$, i. e. $L \subseteq A$. Hence A is the class sum of all those minimal left ideals L of $S \subseteq M$ for which $L \cap A \neq (0)$. Being a sum of left ideals of S, A is itself a left ideal of S.

3. Complete simiplicity of minimal two-sided ide
left and right minimal ideals.

Lemma 3. 1. *Let S be a semigroup without ni*
be a minimal two-sided ideal of S containing at le
one right minimal ideal of S. Then to each minim
tained in M there corresponds at least one minin
contained in M such that $LR = M$ and $RL \neq (0)$.

Proof. ML is a left ideal of S contained in
$ML = (0)$. The second alternative would imply $L^2 \subseteq ML = (0)$, contrary to the assumption that S contains no nilpotent ideal $\neq (0)$. Hence $ML = L$.

LM is a two-sided ideal of S contained in M; hence $LM = M$ or $LM = (0)$. The second alternative is ruled out as above, so that $LM = M$.

Now, by the left-right dual of Theorem 2. 1, M is the class sum of all the minimal right ideals R of S contained in M. Were $LR = (0)$ for every such R, we would conclude $LM = (0)$, contrary to $LM = M$. Hence $LR \neq (0)$ for some minimal right ideal R of S contained in M. But LR is a two-sided ideal of $S \subseteq M$ and $\neq (0)$, whence $LR = M$. Were $RL = (0)$ we would have $L = ML = LRL = (0)$.

The proof of the next sequence of lemmas, culminating in Theorem 3. 1, is the third edition of one due to R. H. Bruck and used by the writer in two previous papers [6]. To avoid repetition, let us state here the hypotheses we assume for them all.

(1) *S is a semigroup without nilpotent ideals.*

(2) *M is a minimal two-sided ideal of S containing at least one left and at least one right minimal ideal of S.*

(3) *L and R are minimal left and right ideals of S contained in M, such that $LR = M$ and $RL \neq (0)$.*

Lemma 3. 2. *If a is any element $\neq 0$ of $R \cap L$, and b is any element of RL, then the equations $ax = b$ and $ya = b$ have solutions x, y in RL.*

Proof. Ma is a left ideal of S contained in L, since $a \varepsilon L$, whence $Ma = L$ or $Ma = (0)$. The latter would imply that the two-element set $(0, a)$ would be a nilpotent left ideal of M, hence of S by Theorem 2. 2. Hence $Ma = L$. aR is a right ideal of S contained in R, since $a \varepsilon R$, and hence $aR = R$ or $aR = (0)$. The latter would imply, using $Ma = L$ and (3), that $M = LR = MaR = (0)$. Hence $aR = R$. Consequently $aRL = RL$, so that $ax = b$ is solvable for x in RL.

Dually, aM is a right ideal of S contained in R. The case $aM = (0)$

$= R$. La is a left ideal of $S \subseteq L$. Were
$= LaM = (0)$. Hence $La = L$, $RLa = RL$,

with zero.

L, whence RL is closed. Since $RL \subseteq R$
RL is a semigroup with zero in which the
eq $= 0)$ are solvable for x and y. Hence, to
show that RL is a group with zero, we need only show that if $a \neq 0$ and $b \neq 0$ then $ab \neq 0$. By (3) there exists an element $c \neq 0$ of RL. Solve $au = c$ for u, then $bx = u$ for x. Then $abx = au = c$. Were $ab = 0$ we would conclude that $c = 0$.

LEMMA 3.4. *Let e be the identity element of the group-with-zero RL. Then $R = eS$, $L = Se$, and $R \cap L = eSe$.*

Proof. Since $e \in R$, eS is a right ideal of S contained in R. $eS \neq (0)$ since it contains $ee = e \neq (0)$. Hence $eS = R$. Similarly $Se = L$. Then $eSe = eS \cap Se = R \cap L$.

LEMMA 3.5. $R \cap L = RL$.

Proof. Clearly we need only show that $R \cap L \subseteq RL$. Let $a \in R \cap L$, and let e be the identity element of RL. If $a = 0$ then $a \in RL$. If $a \neq 0$ then by Lemma 3.2 we can solve $ax = e$ for x in RL. Let x^{-1} be the inverse of x in RL. By Lemma 3.4, $a \in eSe$. Hence $a = ae = axx^{-1} = ex^{-1} = x^{-1} \in RL$.

LEMMA 3.6. *The identity element e of RL is a primitive idempotent.*

Proof. We are to show that the only idempotent elements f of S for which $ef = fe = f$ are $f = e$ and $f = 0$. By Lemma 3.4 such an f belongs to $R \cap L$, hence to RL by Lemma 3.5. RL is a group with zero (Lemma 3.3), and so contains only the idempotents e and 0.

THEOREM 3.1. *Let S be a semigroup without nilpotent ideals. Let M be a minimal two-sided ideal of S containing at least one left and at least one right minimal ideal of S. Then M is a completely simple semigroup.*

Proof. That M is simple follows from Theorem 1.1. By Rees' definition of complete simplicity (*loc. cit.*, p. 393), we must show:

(1) Corresponding to each a in M there exist idempotents e and f in M such that $ea = af = a$.

(2) Every idempotent element $\neq 0$ of M is primitive.

To show (1), let a be any element $\neq 0$ of M (the case $a = 0$ is trivial).

By Theorem 2.1, a belongs to some minimal left ideal L
By Lemma 3.1, there exists a minimal right ideal R
$LR = M$ and $RL \neq (0)$. By Lemma 3.3, RL is a gro
be the identity element thereof. By Lemma 3.4, $L = Sf$.
The dual $ea = a$ is proved similarly.

To show (2), let e be an idempotent element $\neq 0$ of
and its dual, e must belong to a minimal left ideal L
minimal right ideal R of $S \subseteq M$. Then $e = ee \varepsilon LR$, so th[redacted](0). Since LR is a two-sided ideal $\subseteq M$ and $\neq (0)$, $LR = M$. Likewise $RL \neq (0)$ since it contains $ee = e \neq 0$. Hence all three hypotheses prior to Lemma 3.2 are satisfied. By Lemma 3.3, RL is a group with zero. e must be the identity element of RL, and (2) follows from Lemma 3.6.

THEOREM 3.2. *A simple semigroup is completely simple if and only if it contains at least one minimal left and at least one minimal right ideal.*

Proof. A simple semigroup S is a minimal two-sided ideal of itself, by definition of simplicity. Hence if it has the property stated, it is completely simple by Theorem 3.1.

Conversely, let S be completely simple, and let e be a primitive idempotent $\neq 0$ of S. The theorem will follow when we show that Se and eS are minimal left and right ideals of S, respectively.

Suppose L is a left ideal $\neq (0)$ contained in Se, and let a be an element $\neq 0$ of L. Since S is simple, $SaS = S$, and we can solve $xay = e$ for x and y in S. Since $a \varepsilon Se$, $ae = a$. Since $ex \cdot a \cdot eye = eee = e$, we can assume $ex = x$, $ey = ye = y$. Let $f = yxa$. Then $f^2 = yxayxa = yexa = yxa = f$, and $ef = fe = f$. $f \neq 0$ since $y = ye = yxay = fy$. Since e is primitive, $f = e$. Hence $e = yxa \varepsilon Sa \subseteq L$, $Se \subseteq L$, $Se = L$, and Se is minimal. The proof that eS is a minimal right ideal is similar.

The reader will observe that the only assumptions actually used in the second part of the proof were: (1) S is simple, (2) S contains a primitive idempotent e. From these assumptions, then, it follows that S contains minimal left and right ideals, and hence that S is completely simple by Theorem 3. Thus we have an independent proof of a theorem due to Rees that (1) and (2) imply complete simplicity [7].

4. Semigroups in which every two-sided ideal contains minimal left and right ideals. For the three lemmas of this section we assume the following hypotheses:

(1) *S is a semigroup without nilpotent ideals.*

d ideal $\neq (0)$ of S contains at least one left and at
le deal of S.

two-sided ideal $\neq (0)$ of S contains a minimal two-

ny two-sided ideal $\neq (0)$ of S. Let $B[C]$ be the
eft [right] ideals of $S \subseteq A$. By hypothesis (2), B
a Theorem 2.1 and its dual, B and C are two-sided
ideals of S. Applying (2) again to B, we see that B contains at least one minimal right ideal of S, and hence $D = B \cap C \neq (0)$. Clearly D is a two-sided ideal which is the sum of all the minimal left [right] ideals of $S \subseteq D$.

By (2), D contains a minimal left ideal L of S. DL is a left ideal of $S \subseteq L$. $DL = (0)$ would entail $L^2 = (0)$, contrary to (1), whence $DL = L$. LD is a two-sided ideal of $S \subseteq D$. $LD \neq (0)$ since otherwise $L^2 = LDL = (0)$, contrary to (1). Since D is the sum of its minimal right ideals, it must contain a minimal right ideal R of S such that $LR \neq (0)$.

Since LR is a two-sided ideal $\neq (0)$ of S contained in D, and hence in A, the lemma will follow when we show that LR is minimal. Let H be a two-sided ideal $\neq (0)$ of $S \subseteq LR$. Let lr be an element $\neq 0$ of $H (l \varepsilon L, r \varepsilon R)$. Now $Sl = L$. For Sl is a left ideal of $S \subseteq L$, and were $Sl = (0)$, the two-element set $(0, l)$ would be a nilpotent left ideal of S. Similarly $rS = R$. Hence $LR = Sl \cdot rS \subseteq SHS \subseteq H$, whence $H = LR$.

LEMMA 4.2. *Every minimal left ideal of S is contained in some minimal two-sided ideal of S.*

Proof. Let L be a minimal left ideal of S. $L \cup LS$ is a two-sided ideal of S containing L, and we contend that it is minimal. Let B be a two-sided ideal $\neq (0)$ of S contained in it. BL is a left ideal of $S \subseteq L$. $BL = (0)$ would imply $B^2 \subseteq B(L \cup LS) = BL \cup BLS = (0)$, contrary to (1). Hence $BL = L$, and $L \subseteq B$. But then $L \cup LS \subseteq B$, whence equality follows.

LEMMA 4.3. *Every left ideal $\neq (0)$ of S contains a minimal left ideal.*

Proof. Let A be a left ideal $\neq (0)$ of S. By Lemma 4.1, the two-sided ideal AS (evidently $\neq (0)$ since $A^2 \neq 0$) contains a minimal two-sided ideal M of S. Let $m = as$ be an element $\neq 0$ of $M (a \varepsilon A; s \varepsilon S)$. The left ideal $B = (a, Sa)$ consisting of Sa and the single element a, is contained in A, and $Bs = (as, Sas) = (m, Sm) \subseteq M$. Let C be the set of all c in S such that $Bc \subseteq M$. C is clearly a right ideal of S containing the element $s \neq 0$ of S. BC is a two-sided ideal of $S \subseteq M$ and $\neq (0)$ since it contains $m = as \neq 0$. Hence $BC = M$. Now MB is a left ideal of S contained in $M \cap B$, and $\neq (0)$ since otherwise $M^2 = MBC = (0)$. By Theorem 2.1,

M is the sum of minimal left ideals of S, and hence
least one minimal left ideal L of S. The desired res
$A \supseteq B \supseteq MB \supseteq L$.

THEOREM 4.1. *Let S be a semigroup without ni which every two-sided ideal contains at least one left minimal ideal of S. Then every left ideal $L \neq (0)$ o potent element $e \neq 0$. If L is minimal, $L = Se$.*

Proof. By Lemma 4.3, the given left ideal contains a minimal left ideal L of S. By Lemma 4.2, L is contained in a minimal two-sided ideal M of S. By Lemma 3.1, M contains a minimal right ideal R of S such that $LR = M$ and $RL \neq (0)$. Hence the three hypotheses prior to Lemma 3.2 are satisfied. By Lemma 3.3, RL is a group with zero, and the identity element e thereof is a non-zero idempotent belonging to L. By Lemma 3.4, $L = Se$.

An ideal A (left or right) of S is called a nilideal if every element a of A is nilpotent, i. e. $a^n = 0$ for some positive integer n.

COROLLARY 4.1. *Under the hypotheses of Theorem 4.1, S contains no nilideal $\neq (0)$.*

Proof. Suppose A were a left nilideal $\neq (0)$. Then by Theorem 4.1, A would contain an idempotent $e \neq 0$. No power of e is 0, contrary to the assumption that A is a nilideal.

5. Semigroups with kernel $\mathfrak{N}$ and maximal $\mathfrak{N}$-potent ideal $\mathfrak{N}$. Let S be a semigroup in which the intersection $\mathfrak{N}$ of all the two-sided ideals of S is not vacuous. $\mathfrak{N}$ is the "Suschkewitsch kernel" (Rees, *loc. cit.*, p. 392). If S has a zero, $\mathfrak{N} = (0)$. Otherwise $\mathfrak{N}$ is a simple semigroup without zero.

Following Schwarz (*loc. cit.*, p. 39) a left or right ideal A of S is called $\mathfrak{N}$-potent if some power A^n of A is $\subseteq \mathfrak{N}$.

LEMMA 5.1. *The sum of two $\mathfrak{N}$-potent left (or right) ideals is also $\mathfrak{N}$-potent.* (Schwarz, Theorem 38, p. 39).

Proof. Let A and B be $\mathfrak{N}$-potent left ideals, and let $A^m \subseteq \mathfrak{N}$, $B^n \subseteq \mathfrak{N}$. Let $C = A \cup B$. Then $C^{m+n} \subseteq \mathfrak{N}$. For any product of $m + n$ elements of C must contain at least m factors from A or at least n from B; in the first case it belongs to A^m, in the second to B^n.

LEMMA 5.2. *Every $\mathfrak{N}$-potent left (or right) ideal is contained in an $\mathfrak{N}$-potent two-sided ideal.* (Schwarz, Theorem 45, p. 51).

Proof. Let L be an $\mathfrak{N}$-potent left ideal, and let $L^n \subseteq \mathfrak{N}$. Then

$\cup LS$ is $\mathfrak{R}$-potent by Lemma 5. 1, and is a two-

the radical $\mathfrak{R}$ of S to be the sum of all the . $\mathfrak{R}$ is evidently a two-sided ideal of S, and $\mathfrak{R}$-potent left (or right) ideal of S. It need all, however, impose the condition that $\mathfrak{R}$ be alent to requiring that S contain a maximal $\mathfrak{R}$-potent two-sided ideal, which must of course be $\mathfrak{R}$. As an immediate consequence of Lemma 5. 2 we have:

LEMMA 5. 3. *Let S be a semigroup having a kernel $\mathfrak{R}$ and a maximal $\mathfrak{R}$-potent two-sided ideal $\mathfrak{R}$. Then $\mathfrak{R}$ contains every $\mathfrak{R}$-potent left (or right) ideal of S.*

Let A be a two-sided ideal of a semigroup S. Rees (*loc. cit.*, p. 389) defines the difference semigroup $\bar{S} = S - A$ to be that obtained from S by collapsing A into a single zero element $\bar{0}$, while the remaining elements of S retain their identity. Thus there is a one-to-one correspondence $s \leftrightarrow \bar{s}$ between the elements s of S not in A and the non-zero elements $\bar{s}$ of $\bar{S}$, such that $st \leftrightarrow \bar{s}\bar{t}$ if $st \notin A$, and such that $\bar{s}\bar{t} = \bar{0}$ if and only if $st \in A$. This induces a one-to-one correspondence between the class of all left [right, two-sided] ideals B of S containing A and the class of all left [right, two-sided] ideals $\bar{B}$ of $\bar{S}$.

THEOREM 5. 1. *Let S be a semigroup having a kernel $\mathfrak{R}$ and a maximal $\mathfrak{R}$-potent ideal $\mathfrak{R}$. Then the difference semigroup $\bar{S} = S - \mathfrak{R}$ is a semigroup without nilpotent ideals.*

Proof. Suppose $\bar{A}$ is a nilpotent left ideal of $\bar{S}$, and let $\bar{A}^n = (\bar{0})$. Let A be the left ideal of S containing $\mathfrak{R}$ corresponding to $\bar{A}$. Then $A^n \subseteq \mathfrak{R}$. But $\mathfrak{R}^m \subseteq \mathfrak{R}$ for some m, by hypothesis, whence $A^{mn} \subseteq \mathfrak{R}$. Thus A is $\mathfrak{R}$-potent and so $A \subseteq \mathfrak{R}$ by Lemma 5. 3. But this entails $\bar{A} = (\bar{0})$ in the difference semigroup $\bar{S}$.

THEOREM 5. 2. *Let S be a semigroup having a kernel $\mathfrak{R}$ and a maximal $\mathfrak{R}$-potent ideal $\mathfrak{R}$. Assume furthermore that every non-$\mathfrak{R}$-potent two-sided ideal of S contains at least one left and at least one right minimal non-$\mathfrak{R}$-potent ideal of S. Then every left (or right) non-$\mathfrak{R}$-potent ideal of S contains an idempotent element not in $\mathfrak{R}$.*

Proof. By Theorem 5. 1, $\bar{S} = S - \mathfrak{R}$ is a semigroup without nilpotent ideals. The translation to $\bar{S}$ of the stated minimality condition on S is simply

that occurring in the hypothesis of Theorem 4. 1. Let
left ideal of S. Then the corresponding left ideal $\bar{A}$
by Theorem 4. 1, $\bar{A}$ contains an idempotent $\bar{e} \neq \bar{0}$. Th
e in A is then an idempotent not in $\mathfrak{N}$.

We shall say that S satisfies the "descending c
[right, two-sided] ideals" if S does not contain an i
[right, two-sided] ideals A_i $(i = 1, 2, \cdots)$ such that
This is equivalent to the "minimal condition for left [righ[illegible]sided] ideals," that every non-empty set of left [right, two-sided] ideals contain a minimal member. The "ascending chain condition" and the equivalent "maximal condition" are defined in an analogous way.

THEOREM 5.3. *Let S be a semigroup satisfying the descending chain condition both for left and for right ideals, and the ascending chain condition for two-sided ideals. Then S has a kernel $\mathfrak{K}$, which is a completely simple semigroup without zero, and a maximal $\mathfrak{K}$-potent ideal $\mathfrak{N}$. Every left ideal of S not contained in $\mathfrak{N}$ contains an idempotent element not in $\mathfrak{N}$.*

Proof. From the descending chain conditions it is clear that S contains a single minimal two-sided ideal, i. e. that it has a kernel $\mathfrak{K}$, and that $\mathfrak{K}$ is completely simple without zero (Theorem 3. 2, p. 525 of Ref. 1, or Theorem 3. 1 above with the appropriate change in the meaning of "minimal"). From the ascending chain condition it is clear that S contains a maximal $\mathfrak{K}$-potent two-sided ideal $\mathfrak{N}$. Again from the descending chain conditions, the remaining hypothesis of Theorem 5. 2 is satisfied, and the conclusion follows.

COROLLARY 5. 3. *Let S satisfy the conditions of Theorem 5. 3. Then, if S contains no idempotent element outside of $\mathfrak{N}$, S is itself $\mathfrak{K}$-potent. In particular, if S contains at most one idempotent, some power of S is a group.*

Proof. The first part is immediate from Theorem 5. 3, which requires $S = \mathfrak{N}$. The second part follows from the fact that a completely simple semigroup without zero is a sum of groups, and reduces to a group if and only if it contains exactly one idempotent.

This corollary is very much like Schwarz's Theorem 42 (p. 46). Schwarz assumes that every element of S has finite order, but requires the descending chain condition only for left ideals. He notes that the finite order requirement can be dropped if S is commutative, in which case there is of course no difference between left and right ideals.

The ascending chain condition was used only to insure that $\mathfrak{N}$ be $\mathfrak{K}$-

ter condition may replace the former in the hypotheses
corollary. The new corollary then bears the same
Theorem 46 (p. 52) that the old one did to his

remark on a second possible definition (and there
the radical of a semigroup. Call a left [right]
ia left [right] nilideal if every element a of A is $\mathfrak{R}$-potent, i. e. $a^n \in \mathfrak{R}$ for some n. The sum $\mathfrak{R}^*$ of all two-sided nil-ideals is a two-sided nilideal containing all left and all right nilideals. Evidently $\mathfrak{R}^* \supseteq \mathfrak{R}$. But if S satisfies the hypotheses of Theorem 5.2 (or 5.3) then $\mathfrak{R}^* = \mathfrak{R}$. For clearly no nilideal can contain an idempotent outside of $\mathfrak{R}$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

1. A. H. Clifford, "Semigroups containing minimal ideals," *American Journal of Mathematics*, vol. 70 (1948), pp. 521-526. This paper was in turn an extension to the infinite case of certain results of A. Suschkewitsch, "Über die endlichen Gruppen ohne das Gesetz der eindeutigen Umkehrbarkeit," *Mathematische Annalen*, vol. 99 (1928), pp. 30-50.
2. D. Rees, "On semi-groups," *Proceedings of the Cambridge Philosophical Society*, vol. 36 (1940), pp. 387-400; the definition is given on p. 393.
3. In a forthcoming paper, R. P. Rich points out that the condition that S contain no nilpotent ideal $\neq (0)$ can be replaced by the much weaker and more appropriate condition $M^2 \neq (0)$, throughout the first three sections of the present paper.
4. R. Baer and F. Levi, "Vollständige irreduzibele Systeme von Gruppenaxiomen," *Sitzungsberichte der Heidelberger Akademie der Wissenschaften, Mathematisch-naturwissenschaftliche Klasse, Beiträge zur Algebra*, No. 18 (1932), 12 pp.; the example is given on p. 7.
5. Stefan Schwarz, "Zur Theorie der Halbgruppen," *Sborník prác Prírodovedeckej fakulty Slovenskej univerzity v Bratislave*, No. 6 (1943), 64 pp. (Slovakian, German summary); *Mathematical Reviews*, vol. 10 (1949), p. 12.
6. A. H. Clifford and D. D. Miller, "Semigroups having zeroid elements," *American Journal of Mathematics*, vol. 70 (1948), pp. 117-125. Bruck's proof is used in § 2 of this paper and again in § 3 of ref. 1 above.
7. D. Rees, "Note on semi-groups," *Proceedings of the Cambridge Philosophical Society*, vol. 37 (1941), pp. 434-435.

THE RADIUS OF UNIVALENCE OF AN ANALYTIC FUNCTION.*

By Zeev Nehari.

A family of analytic functions $f(z)$ which is no
domain D and satisfies $|f'(\xi)| \geq C > 0$ at a point
univalence about ξ, i. e. there exists a positive number r such that all functions of the family are schlicht in the interior of the circle $|z - \xi| < r$. While the existence of a circle of univalence follows easily from the theory of normal families, the determination of the "exact" radius of univalence of a given family of functions may be a very difficult task. In the case of the family of bounded functions in the unit circle, the exact radius of univalence was found by Landau [3].

In the present paper, we shall determine the exact radii of univalence of two families of functions which are defined in an arbitrary bounded domain D of finite connectivity.

Let S denote the family of functions $f(z)$ for which the function $\log [f(z) - f(\xi)]/[z - \xi]$ is regular and single-valued in D, where $\xi \varepsilon D$ is a given fixed point of D. For the sake of convenience we shall further assume that $\xi = 0$. $f(z)$ being in S, this implies that also $f(0) = 0$. Obviously, this assumption is no restriction on the generality of our considerations.

We begin with the following theorem:

Theorem I. *Let $f(z)$ be in S and let $M = \underset{z \varepsilon D}{\text{l. u. b.}} |\log [f(z)/z]|$; then $f(z)$ is schlicht in the largest circle about the origin all of whose points satisfy*

$$|z| K(z, z) \leq 2\pi/M. \tag{1}$$

Here, $K(z, \xi)$ denotes the Szegö kernel function [2, 4] *of D. This value of the radius of univalence is the best possible.*

Proof. By hypothesis, $\log [f(z)/z]$ is regular and single-valued in D. We therefore have, by the residue theorem,

$$f'(z)/f(z) - 1/z = \frac{1}{2\pi i} \int_\Gamma \log(f(z)/x) q(x)\, dx, \tag{2}$$

---

* Received August 18, 1948.

where $q(z)$ is an arbitrary single-valued function in D which is regular in
; $x = z$ where it has the principal part $(x - z)^{-2}$.
on be permissible, we assume further that the boun-
ooth curves and that $f(x)$ is continuous on Γ; once
h assumptions can easily be disposed of. Indeed,
nded, we may approximate D by a sequence of
$D_n \subset D$, $D_n \subset D_{n+1}$, $\lim_{n=\infty} D_n = D$, and whose boun-
we replace D by D_n, the additional assumptions
under which we prove Theorem I are satisfied. The general result then follows by letting $n \rightarrow \infty$ and observing that l. u. b. $|\log z^{-1}f(z)| \leq M$ for $z \varepsilon D_n$, and that the Szegö kernel function $K(z,z)$ is a continuous domain function.

The particular function $q(x)$ in (2) we shall choose is connected with the function $F(x) = F(x,z)$, introduced by Ahlfors [1], which solves the following extremal problem: Among all functions $g(x)$ which are regular and single-valued in D and satisfy there $|g(x)| \leq 1$, $g(z) = 0$, to find the function which maximizes $|g'(z)|$. If n is the connectivity of D, it was shown by Ahlfors that $F(x)$ maps D onto the n-times covered unit circle. More recently, it was shown by Garabedian [2] that there exists a function $q(x)$ which is regular in $D + \Gamma$ except at the point $x = z$, where its principal part is $(x - z)^{-2}$, and which has the property

$$(1/i)F(x)q(x)\,dx > 0, \qquad x \varepsilon \Gamma. \tag{3}$$

We now identify $q(x)$ in (2) with this particular function. By (2), we have

$$I = |f'(z)/f(z) - 1/z| \leq \tfrac{1}{2}M \int_\Gamma |q(x)\,dx|.$$

Since, on Γ, $|F(x)| \equiv 1$, this may also be written

$$I \leq \frac{1}{2\pi} M \int_\Gamma |F(x)q(x)\,dx| = \frac{1}{2\pi i} M \int_\Gamma F(x)q(x)\,dx,$$

the last identity following from (3). Using the residue theorem, we therefore obtain $|f'(z)/f(x) - 1/z| \leq MF'(z)$ or

$$|zf'(z)/f(z) - 1| \leq M\,|z|\,F'(z). \tag{4}$$

A sufficient condition for $f(z)$ to be schlicht in a circle C about the origin is

$$\mathrm{Re}\{zf'(z)/f(z)\} \geq 0 \tag{5}$$

in C; (5) even assures that C be mapped by $f(z)$ onto a star-like domain. (5) is obviously satisfied if $|zf'(z)/f(z)-1| \leq 1$. In view of (4), a circle about the origin all of whose points satisfy

$$M\,|z|\,F'(z) \leq 1 \tag{6}$$

will therefore be mapped by $f(z)$ onto a schlicht sta

It was shown by Garabedian [2] that $2\pi F'(z) =$
denotes the kernel

$$K(z,\xi)=\sum_{\nu=1}^{\infty}\phi_\nu(z)\overline{\phi_\nu(\xi)}$$

of a complete system of regular functions in D which is ortho-normalized by the conditions

$$\int_\Gamma \phi_\nu(z)\overline{\phi_\mu(z)}\,ds=\delta_{\mu\nu}, \qquad ds=|dz|.$$

(6) may therefore also be written $|z|\,K(z,z)\leq 2\pi/M$, which is identical with (1). This shows that the circle defined in the statement of Theorem I is indeed mapped by $f(z)$ onto a schlicht domain.

In order to show that this is the best possible lower bound for the radius of univalence, we consider the function

$$f_0(z)=z\exp\{-Me^{-i\gamma}F(z,re^{i\gamma})\},$$

where $z=re^{i\gamma}$ is the point nearest to the origin among those points z for which equality holds in (1). Obviously,

$$\operatorname*{l.u.b.}_{z\epsilon D}|\log f_0(z)/z|=M.$$

Setting $f_0(z)=f(z)$ in (2), we obtain

$$f'_0(re^{i\gamma})/f_0(re^{i\gamma})-1/(re^{i\gamma})=-\tfrac{1}{2\pi}K(re^{i\gamma},re^{i\gamma})Me^{-i\gamma}.$$

Since for $z=re^{i\gamma}$ we have equality in (1), we have therefore

$$f'_0(re^{i\gamma})/f_0(re^{i\gamma})-1/(re^{i\gamma})=-1/(re^{i\gamma}).$$

whence $f'_0(re^{i\gamma})=0$. This shows that $f_0(z)$ cannot be schlicht in a circle about the origin with a radius greater than r.

As an example we consider the case where D is the unit circle $|z|\leq 1$. In this case, $K(z,z)=2\pi(1-|z|^2)^{-1}$ [4]. By Theorem I, the radius of univalence of the family of functions for which $|\log[f(z)/z]|\leq M$ in $|z|<1$ is therefore $r=\frac{1}{2}\{(M^2+4)^{\frac{1}{2}}-M\}$. The exact value of this radius of univalence is attained by the function

$$f_0 = z \exp\{-M([z-r]/[1-rz])\}.$$

The next theorem gives the circle of univalence of a function $f(z)$ in S, tive number N satisfying

$$(7 \qquad e^{-N} < |f(z)/z| < e^{N}$$

is heorem I can be used as an N but not vice versa, th esult.

THEOREM II. *Let $f(z)$ be in S and let N be defined by* (7); *then $f(z)$ is schlicht in the largest circle about the origin all of whose points satisfy*

$$(8) \qquad |z|\,K(z,z) \leq \pi^2/(2N),$$

where $K(z,\zeta)$ is again the Szegö kernel. This lower bound for the radius of univalence is the best possible.

Proof. We start again with the identity (2), but we now identify $q(x)$ with a different function. If $F(x) = F(x,z)$ denotes again the same function as above, we find, in view of $|F(x)| \equiv 1$, $x \in \Gamma$, that

$$\theta F(x)/\{1+\theta^2F^2(x)\} \qquad (|\theta| = 1)$$

is real on Γ. In view of (3), the differential

$$(9) \qquad (1/i)\,q_1(x)\,dx = (1/i\theta)\,q(x)\,[1+\theta^2F^2(x)]\,dx$$

will also be real on Γ. The principal part of $q_1(x)$ at $x = z$ is $\theta^{-1}(x-z)^{-2}$. We may therefore use $q_1(x)$ instead of $q(x)$ in (2). This yields

$$\theta^{-1}(f'(z)/f(z) - 1/z) = \tfrac{1}{2\pi i}\theta^{-1}\int_\Gamma \log(f(x)/x)\,[1+\theta^2F^2(x)]\,q(x)\,dx.$$

Taking real parts, we obtain, in view of the reality of the differential (9),

$$Re\{1/\theta(f'(z)/f(z) - 1/z)\} = \tfrac{1}{2\pi i}\theta^{-1}\int_\Gamma \log|f(x)/x|\,[1+\theta^2F^2(x)]\,q(x)\,dx,$$

whence

$$(10) \qquad |Re\{\theta^{-1}(f'(z)/f(z) - 1/z)\}| \leq \tfrac{1}{2\pi}N\int_\Gamma |[1+\theta^2F^2(x)]\,q(x)\,dx|.$$

We cannot drop now the absolute value signs at the right-hand side and use the residue theorem, since the differential (9), while real, is not of constant sign. This difficulty can, however, be overcome in the following way. Writing

$$q_1(x)\,dx/i = [F(x)\,q(x)\,dx/i]\,[1+\theta^2F^2(x)]/[\theta F(x)],$$

we see that $q_1(x)\,dx/i$ is positive for $-\pi/2 < \arg \theta F(x) < \pi/2$ and negative for $\pi/2 < \arg \theta F(x) < 3\pi/2$. Since, if

$$G_\theta(x) = [1 + i\theta F(x)]/[1 - i\theta F(a$$

$2\pi^{-1} \arg G_\theta(x)$ is equal to 1 for $-\pi/2 < \arg \theta F(x) <$
for $\pi/2 \arg \theta F(x) < 3\pi/2$, we have

$$\begin{aligned} |\, q_1(x)\,dx/i \,| &= (\tfrac{1}{2}\pi i\theta)^{-1} q(x)\,[1 + \theta^2 F^2(x)] \arg G_\theta(x)\,dx \\ &= (\tfrac{1}{2}\pi i\theta)^{-1} q(x)\,[1 + \theta^2 F^2(x)] \operatorname{Im} \log G_\theta(x)\,dx \\ &= \operatorname{Im}\{ (\tfrac{1}{2}\pi i\theta)^{-1} q(x)\,[1 + \theta^2 F^2(x)] \log G_\theta(x)\,dx \}. \end{aligned}$$

Using (10), we thus obtain

$$|\, \operatorname{Re} \theta^{-1}\{f'(z)/f(z) - 1/z\} \,| \leq 2\pi^{-1}N,\ \operatorname{Im}\{ \tfrac{1}{2\pi i} \theta^{-1} \int_\Gamma q(x)\,[1 + \theta^2 F^2(x)] \log G_\theta(x)\,dx.$$

While the integrand has singularities on Γ, it is continuous there. We may therefore evaluate the integral on the right-hand side by the residue theorem. This yields

$$|\, \operatorname{Re} \theta^{-1}(f'(z)/f(z)) - 1/z \,| \leq 2\pi^{-1}N, \qquad \operatorname{Im}\{2iF'(z)\} = 4\pi^{-1}NF'(z),$$

or, because of the arbitrariness of the argument of θ,

$$|\, zf'(z)/f(z) - 1 \,| \leq (4\,|\,z\,|\,N/\pi)F'(z) = (2N/\pi^2)\,|\,z\,|\,K(z,z).$$

For values z which satisfy (8) we shall therefore have

$$|\, zf'(z)/f(z) - 1 \,| \leq 1.$$

Since such values satisfy a fortiori the inequality (5), a circle about the origin all of whose points z are subject to (8) will be mapped by $f(z)$ onto a schlicht star-like domain.

In order to show that this lower bound for the radius of univalence is the best possible, we denote by $re^{i\gamma}$ the point nearest to the origin for which equality holds in (8) and consider the function

$$f_0(z) = z([1 - ie^{-i\gamma}F(z)]/[1 + ie^{-i\gamma}F(z)])^{2N/\pi i}, \qquad F(z) \equiv F(z, re^{i\gamma}).$$

Clearly,

$$e^{-N} \leq |\, f_0(z)/z \,| \leq e^{N}, \qquad z \in D.$$

We have

$$1/r - re^{i\gamma}f'_0(re^{i\gamma})/f_0(re^{i\gamma})$$

$$= N \text{[illegible]} 1 - ie^{-i\gamma}F(x)]/[1 + ie^{-i\gamma}F(x)])(1 + e^{-2i\gamma}F^2(x))q(x)dx$$

$$= \text{[illegible]} = (2/\pi^2)NK(re^{i\gamma}, re^{i\gamma}) = 1/r,$$

[illegible] from the fact that, for $z = re^{i\gamma}$, equality holds in (8). [illegible] which shows that $f_0(z)$ cannot be schlicht in a circle [illegible] us larger than r. This completes the proof of Theorem [illegible]

As an illustration, we again consider the case where D is the unit circle. [illegible], in this case, $K(z, z) = 2\pi(1 - |z|^2)^{-1}$, we have the result: If $z^{-1}f(z) \neq 0$ in $|z| < 1$ and $e^{-N} \leq |z|^{-1} |f(z)| \leq e^N$, then $f(z)$ is schlicht in the circle $|z| < r$ with

$$r = \pi^{-1}[(4N^2 + \pi^2)^{\frac{1}{2}} - 2N].$$

The exact value of this bound is attained by the function

$$f_0(z) = z([1 + ir - iz(1 - ir)]/[1 - ir + iz(1 + ir)])^{2N/\pi i}.$$

It is worthy of note that, as shown by a trivial modification of the proof, the lower bound for the radius of univalence given by Theorem II remains also valid if the condition $|\operatorname{Re}\{\log z^{-1}f(z)\}| \leq N$ is replaced by

$$|\operatorname{Re}\{e^{i\beta}\log[f(z)/z]\}| \leq N,$$

where β $(0 \leq \beta < 2\pi)$ is arbitrary. In particular, the constant N in Theorem II may be so chosen as to satisfy

$$|\arg f(z) - \arg z| \leq N, \qquad z \varepsilon D.$$

By a slight modification of the methods of proof used above it is possible to obtain two sharp estimates for the radius of convexity of the family T of functions for which $\log f'(z)$ is regular and single-valued in D. Here, the radius of convexity of a function $f(z)$ about $z = 0$ is defined as the radius of the laregst circle about the origin which is mapped by $f(z)$ onto a convex schlicht domain. We have the following two results.

THEOREM III. *Let $f(z)$ be in T and let*

$$|\log f'(z)| \leq M, \qquad z \varepsilon D;$$

let further r denote the radius of the largest circle about the origin all of whose points satisfy

$$|z|\,K(z,z) \leq 2\pi/M;$$

then the circle $|z| < r$ *is mapped by* $f(z)$ *onto a convex schlicht domain. This lower bound for the radius of convexity is the best*

THEOREM IV. *Let* $f(z)$ *be in* T *and* N *such that*

$$e^{-N} \leq |f'(z)| \leq e^{N} \tag{11}$$

or

$$|\arg f'(z)| \leq N \tag{12}$$

is satisfied. If r *denotes the radius of the largest circle about the origin all of whose points satisfy*

$$|z|\,K(z,z) \leq \tfrac{1}{2}\pi^2/N,$$

then the circle $|z| < r$ *is mapped by* $f(z)$ *onto a convex schlicht domain. This bound is again the best possible, irrespective of whether* N *is defined by* (11) *or* (12).

We have

$$f''(z)/f'(z) = \frac{1}{2\pi i}\int_{\Gamma} \log f'(x)\,q(x)\,dx, \tag{13}$$

where $q(x)$ has the same meaning as in (2). Hence,

$$\begin{aligned} |f''(z)/f'(z)| &\leq \frac{1}{2\pi} M \int_{\Gamma} |q(x)\,dx| = \frac{1}{2\pi} M \int_{\Gamma} |F(x)q(x)\,dx| \\ &= \frac{1}{2\pi i} M \int_{\Gamma} F(x)q(x)\,dx = MF'(z) = \frac{1}{2\pi} MK(z,z), \end{aligned} \tag{14}$$

where $F(x)$ again denotes the Ahlfors extremal function.

A necessary and sufficient condition for a circle about the origin to be mapped onto a convex schlicht domain is

$$1 + \operatorname{Re}\{zf''(z)/f'(z)\} \geq 0.$$

This condition is certainly satisfied if

$$|zf''(z)/f'(z)| \leq 1.$$

In view of (14), any circle about the origin all of whose points satisfy

$$|z|\,K(z,z) \leq 2\pi/M \tag{15}$$

will therefore be a circle of convexity of $f(z)$.

In order to show that this bound is the best possible, we denote by $re^{i\gamma}$ the nearest point to the origin for which equality is attained in (15), and consider the function

$$(z) = \int_0^z \exp\{-Me^{-i\gamma}F(z, re^{i\gamma})\}dz.$$

Substituting $f_0(z)$ for $f(z)$ in (13), it is readily found that

$$f''(re^{i\gamma})/f'(re^{i\gamma}) = -\tfrac{1}{2\pi}Me^{i\gamma}K(re^{i\gamma}, re^{i\gamma}) = -e^{-i\gamma}/r,$$

the last step following from the fact that, for $z = re^{i\gamma}$, equality holds in (14). Hence,

$$1 + re^{i\gamma}f''(re^{i\gamma})/f'(re^{i\gamma}) = 0,$$

and this shows that a circle of radius larger than r cannot be mapped by $f(z)$ onto a convex domain, since in the vicinity of $re^{i\gamma}$ there must be points z for which

$$1 + \operatorname{Re}\{zf''(z)/f'(z)\} < 0.$$

The proof of Theorem IV and the construction of the extremal functions follow in a similar way by making the necessary modifications in the proof of Theorem IV. We omit them here, since they present no new features.

HARVARD UNIVERSITY AND
WASHINGTON UNIVERSITY.

REFERENCES.

1. L. V. Ahlfors, "Bounded analytic functions," *Duke Mathematical Journal*, vol. 14 (1947), pp. 1-11.
2. P. Garabedian, *Schwarz's lemma and the Szegö kernel function*, Thesis, Harvard University, 1948.
3. E. Landau, "Der Picard-Schottkysche Satz und die Blochsche Konstante," *Sitzungsberichte, Akademie der Wissenschaften*, Berlin, *Physikalische-Mathematische Klasse*, 1926, pp. 467-474.
4. G. Szegö, "Ueber orthogonale Polynome, die zu einer gegebenen Kurve der komplexen Ebene gehoeren," *Mathematische Zeitschrift*, vol. 9 (1921), pp. 218-270.

ON LINEAR ASYMPTOTIC EQUILIBRIA.*

By AUREL WINTNER.

For large positive t, say for

$$0 \leqq t < \infty, \tag{1}$$

let (a_{ik}), where $1 \leqq i \leqq n$, $1 \leqq k \leqq n$ and $a_{ik} = a_{ik}(t)$, be a matrix of n^2 continuous functions. Conditions will be considered under which this matrix has the following property: All n components, $x_1(t), \cdots, x_n(t)$, of every solution of the linear differential system

$$x_i' = \sum_{k=1}^{n} a_{ik}(t) x_k \tag{2}$$

tend, as $t \rightarrow \infty$, to finite limits and, if $c_1, \cdots, c_n$ are arbitrary constants, (2) has a solution corresponding to which the n limits $x_i(\infty)$ become the respective given values c_i.

If this is the case, let (2) be called of type (*). The simplest example of a system of type (*) is the trivial system $z_i' = 0$, with $z_1(t) = c_1, \cdots, z_n(t) = c_n$ as its general solution. To say that (2) is of type (*) means that the solutions of (2) are in asymptotic one-to-one correspondence with the solutions of the trivial system, where $(a_{ik}) \equiv (0)$.

(I) *Let each of the n^2 continuous functions $a_{ik}(t)$ satisfy the following three conditions: The integral*

$$\int^{\infty} a_{ik}(t)\,dt \text{ is convergent in the sense } \int^{\infty} = \lim_{T\rightarrow\infty} \int^{T} \tag{3}$$

and, when considered as a function of the lower limit of integration, behaves so as to satisfy

$$\int^{\infty} \left| \int_{s}^{\infty} a_{ik}(t)\,dt \right| ds < \infty; \tag{4}$$

finally,

$$\limsup_{t\rightarrow\infty} | a_{ik}(t) | < \infty. \tag{5}$$

Then (2) *is of type* (*).

---

* Received October 18, 1948.

Easy examples show that (3), (4), (5) together fail to imply

$$\int^{\infty} | a_{ik}(t) | \, dt < \infty. \tag{6}$$

On the [illegible] while (6) implies (3), neither (4) nor (5) follows from (5). Hence, (1) neither contains, nor is contained in, the following criterion:

(II) *Let each of the n^2 continuous functions $a_{ik}(t)$ be such as to satisfy* (6). *Then* (2) *is of type* (*).

Still another criterion runs as follows:

(III) *Let each of the n^2 continuous functions $a_{ik}(t)$ satisfy the following three conditions*: (3),

$$\int^{\infty} | \int_{s}^{\infty} a_{ik}(t) \, dt |^{p} \, ds < \infty \tag{4 bis}$$

and

$$\int^{\infty} | a_{ik}(t) |^{q} \, dt < \infty, \tag{5 bis}$$

where $p > 1$ is a fixed index and $q = p/(p-1)$. Then (2) *is of type* (*).

Clearly, (I) and (II) can be interpreted as limiting cases, $p = 1$ and $p = \infty$, of (III).

(II) is known.[1] The proof of (I) will be more elaborate. It will be given in a form which will make it clear that the proof of (III) is exactly the same as that of (I), except that the applications of the "first mean-value theorem," to be based on (4) and (5), must be replaced by corresponding applications of Hölder's inequality.

The proof of (I) proceeds as follows:

If A denotes the matrix (a_{ik}), and x the column vector in which the i-th component is x_i, then (2) can be written as

$$x' = A(t)x. \tag{7}$$

---

[1] The result goes back to Bôcher; cf. O. Dunkel, "Regular singular points of a system of homogeneous linear differential equations of the first order," *Proceedings of the American Academy of Arts and Sciences*, vol. 38 (1902), pp. 341-370, where the independent variable corresponds to e^{-t}, if t is the independent variable used above. For a direct proof, cf. A. Wintner, "Small perturbations," *American Journal of Mathematics*, vol. 67 (1945), pp. 417-430 (more particularly p. 428).

Conditions (3), (4), (5) respectively mean that, on the half-line (1),

$$(8)\qquad B(t)=\int_t^\infty A(s)\,ds, \text{ where } \int_t^\infty=\lim_{T\to\infty}\int_t^T$$

defines a matrix function, $B(t)$; that

$$(9)\qquad b(t)=\int_t^\infty |B(s)|\,ds$$

defines there a scalar function, $b(t)\neq\infty$; and that there exists a constant, β, satisfying

$$(10)\qquad |A(t)|\leqq\beta.$$

In (9) and (10), the sign of absolute value is meant in the following sense: If C is a matrix, $|C|$ denotes the maximum of $|Cv|$ when $|v|=1$, where $|Cv|$ denotes the length of the vector Cv into which C transforms a vector v of length $|v|=1$. Thus $|C_1+C_2|\leqq|C_1|+|C_2|$ and $|C_1C_2|\leqq|C_1|\,|C_2|$.

Let $y(t)$ be any vector function which is continuous on (1) and satisfies

$$(11)\qquad \limsup_{t\to\infty}|y(t)|<\infty.$$

Then it is seen from the convergence of the integral (8) that (11) remains true if $y(t)$ is replaced by $B(t)y(t)$. Similarly, (11), (10) and the convergence of (9) imply that the integral

$$\int_t^\infty B(s)A(s)y(s)\,ds$$

is (absolutely) convergent and, when denoted by $z(t)$, is such that (11) remains true if $y(t)$ is replaced by $z(t)$.

In view of these two facts, it is possible to define on (1) a sequence of continuous vector functions $x^0(t), x^1(t), \cdots$ by the recursion formula

$$(12)\qquad x^{m+1}(t)=c-B(t)x^m(t)-\int_t^\infty B(s)A(s)x^m(s)\,ds,$$

where c is an arbitrarily chosen constant vector and

$$(13)\qquad x^0(t)\equiv c.$$

If m in (12) is replaced by $m-1$ and the resulting relation is subtracted from (12), it is seen that

$$\mu_{m+1}(t) \leq |B(t)|\,\mu_m(t) + \int_t^\infty |B(s)|\,|A(s)|\,\mu_m(s)\,ds,$$

where $\mu_m(t) = |x^m(t) - x^{m-1}(t)|$. In particular, since the function

$$\lambda_m(t) = \operatorname*{fin\,sup}_{t \leqq s < \infty} |x^m(s) - x^{m-1}(s)| \tag{14}$$

is a non-increasing majorant of $\mu_m(t)$,

$$\mu_{m+1}(t) \leqq |B(t)|\,\lambda_m(t) + \lambda_m(t) \int_t^\infty |B(s)|\,|A(s)|\,ds.$$

Hence, by (10) and (9),

$$\mu_{m+1}(t) \leqq (|B(t)| + b(t)\beta)\lambda_m(t).$$

Since β is a constant, the convergence of (8) and (9) implies that, from a certain t onward, $|B(t)| + b(t)\beta < \frac{1}{2}$. It follows therefore from the last formula line that, if t_0 is large enough and

$$t_0 \leqq t < \infty, \tag{15}$$

then $\lambda_{m+1}(t) < \frac{1}{2}\lambda_m(t)$, hence $\lambda_{m+1}(t) < \lambda_1(t)/2^m$. Consequently, if it is ascertained that $\lambda_1(t) \rightarrow 0$ as $t \rightarrow \infty$, it will follow that, on the half-line (15), the inequality

$$\lambda_m(t) < \epsilon(t)/2^m \tag{16}$$

holds for a function $\epsilon(t)$ which does not depend on m and tends to 0 as $t \rightarrow \infty$. But (13), the case $m = 0$ of (12) and the case $m = 1$ of (14) show that $\lambda_1(t) \rightarrow 0$ is equivalent to

$$-B(t)c - \int_t^\infty B(s)A(s)c\,ds \rightarrow 0, \qquad (t \rightarrow \infty),$$

and the truth of the latter limit relation, in which c is a constant, is clear from (10) and from the convergence of both integrals (8), (9). This proves that $\epsilon(t) \rightarrow 0$ in (16).

It follows from (14) and (16) that, on the half-line (15), a vector function $x(t)$ can be defined by placing

$$x(t) = x^0(t) + \sum_{m=1}^{\infty} \{x^{m+1}(t) - x^m(t)\}.$$

The convergence of this series is uniform on (15); so that, since every $x^m(t)$ is continuous, $x(t)$ is. Furthermore, by (13),

$$x(t) \rightarrow c \text{ as } t \rightarrow \infty, \tag{17}$$

since $\sum_{m=1}^{\infty} \epsilon(t)/2^m = \epsilon(t) \rightarrow 0$. Finally, since $x(t) = \lim_{m \rightarrow \infty} x^m(t)$,

$$x(t) = c - B(t)x(t) - \int_t^{\infty} B(s)A(s)x(s)ds. \tag{18}$$

In fact, (10) and the convergence of (9) imply that the limit process, $m \rightarrow \infty$, can be carried out beneath the integral sign of (12), since, in view of (14) and (16), the functions $x^m(t)$ are uniformly bounded on the half-line (15).

It will be shown that, if t is large enough, the integral equation (18) implies the differential equation (7). If this is granted, it will follow that (7) has a solution satisfying (17), where c is any given constant vector. This will complete the proof of (I).

In fact, let ${}^ie$ be the column vector corresponding to which the matrix $({}^1e, \cdots, {}^me)$ becomes the unit matrix, and let ${}^ix(t)$ denote the solution $x(t)$ which the above construction supplies when $c = {}^ie$. Then $\det({}^1x, \cdots, {}^nx)$ tends to 1 as $t \rightarrow \infty$. Hence, the n solution vectors ${}^ix(t)$ are linearly independent. Consequently, if $x(t)$ is any solution of (7), there exists a unique set of n scalar constants γ_i satisfying

$$x(t) = \sum_{i=1}^{n} {}^ix(t)\gamma_i.$$

It follows therefore from $\lim {}^ix(t) = {}^ie$ that $\lim x(t)$ exists for every solution $x(t)$, the vector c in (17) being

$$c = \sum_{i=1}^{n} {}^ie\gamma_i.$$

Accordingly, (7) is of type (*), as claimed in (I).

In order to prove that (18) is just a transcription of (7), put

$$C(t) = E + B(t) \tag{19}$$

where E is the unit matrix. Then the convergence of the integral (8) implies that $C(t) \rightarrow E$ as $t \rightarrow \infty$. Hence, $\det C(t) \neq 0$ on the half-line (15), if t_0 is large enough. Thus $C^{-1}(t)$ exists on (15).

According to (18) and (19),

$$C(t)x(t) = c - \int_t^\infty B(s)A(s)x(s)\,ds. \tag{20}$$

Si (8) are continuous functions, the vector on the right of (20) has a first derivative. It follows therefore from (20) that the same is true of the vector $C(t)x(t)$. Consequently, the same is true of the vector $x(t)$ if it is true of the matrix $C^{-1}(t)$. But $C^{-1}(t)$ has a derivative if $C(t)$ does; in fact,

$$(C^{-1})' = - C^{-1}C'C^{-1} \text{ if } \det C \neq 0, C = C(t).$$

Since (19) and (8), where $A(t)$ is continuous, imply the existence of $C'(t)$, the existence of $x'(t)$ follows.

Consequently, (18) can be differentiated. This gives

$$x'(t) = - B'(t)x(t) - B(t)x'(t) + B(t)A(t)x(t).$$

But $B'(t) = - A(t)$, by (8). Hence, if the first term on the left of the last formula line is moved to the right, it is seen that $w(t) = - B(t)w(t)$, where $w(t)$ is an abbreviation for $x'(t) - A(t)x(t)$. Accordingly, $C(t)w(t) = 0$, by (19). It follows therefore from $\det C(t) \neq 0$ that $w(t) = 0$. In view of the definition of $w(t)$, this proves that $x(t)$ is a solution of (7).

THE JOHNS HOPKINS UNIVERSITY.

ON THE CLASSICAL EXISTENCE THEOREM OF LINEAR DIFFERENTIAL EQUATIONS.*

By PHILIP HARTMAN and AUREL WINTNER

---

1. In the classical existence theorem of systems of linear differential equations

$$dw_i/dz = \sum_{k=1}^{n} g_{ik}(z)w_k, \qquad (i = 1, \cdots, n),$$

the coefficient functions are supposed to be regular power series, convergent in a circle about $z = 0$. If $z = e^{-s}$, the system appears in the form

$$(1) \qquad dw_i/ds = \sum_{k=1}^{n} g_{ik}(s)w_k, \qquad (i = 1, \cdots, n),$$

where the coefficient functions are convergent series in positive powers of e^{-s}, power series $a_1e^{-s} + a_2e^{-2s} + \cdots$ *having no constant terms* (and convergent for large $\Re s$). It was shown in [2] that, due to this circumstance, the classical existence theorem can be extended to the case in which the coefficient functions of (1) are absolutely convergent Laplace-Stieltjes integrals with a *positive* lower limit of integration,

$$(2) \qquad \int_{\lambda_0}^{\infty} e^{-s\lambda} d\alpha(\lambda), \qquad \lambda_0 > 0,$$

where $s = \sigma + it$ and $\sigma >$ const.

The present note deals with another extension, namely, with the case in which the coefficient functions are uniformly almost-periodic (regular) functions on a half-plane $\sigma >$ const., with Fourier expansions of the form

$$(3) \qquad \sum_{m} a_m e^{-\lambda_m s}, \qquad \lambda_m \geqq \lambda_0 > 0,$$

which need not have a half-plane of convergence (and still less, as in (2), a half-plane of absolute convergence). While both extensions contain the classical theorem, neither can be deduced from the other and it seems to be hard to formulate *one* theorem containing *both* extensions.

---

* Received November 5, 1948.

It may be mentioned that the analogue of existence theorems in question is not likely to be true in case of integrals (2) having half-planes of convergence but no half-planes of absolute convergence.

on $f(s)$ be called uniformly almost-periodic in the closed half- on the one hand, $f(s)$ is regular and (in the usual sense) uniformly almost-periodic on the open half-plane $\sigma > 0$, and, on the other hand, $f(s)$ goes over on the line $\sigma = 0$ into a continuous boundary function, $f(it)$, which is uniformly almost-periodic (for $-\infty < t < \infty$), while the uniform almost-periodicity of the functions $f(\sigma + it)$ of the real variable t is uniform in σ (as $\sigma \rightarrow +0$).

In this terminology, the theorem in question can be formulated as follows:

THEOREM. *In the half-plane $\sigma \geqq 0$, let the n^2 coefficient functions of* (1) *be uniformly almost-periodic, with Fourier expansions,*

$$f_{ik}(s) \sim \sum_m a_m{}^{ik} e^{-\lambda_m s}, \tag{4}$$

in which the exponents, λ_m, have a positive lower bound. Then every component, $w_i(s)$, of every solution, $(w_1, \cdots, w_n)$, of (1) *is uniformly almost-periodic in the half-plane $\sigma \geqq 0$. Furthermore, the n integration constants, c_i, specifying a solution of* (1) *can be chosen to be the constant terms of the Fourier expansions of $w_i(s)$, and any set of n assigned mean-values, c_i, determines a solution $(w_1, \cdots, w_n)$ uniquely. Finally, the non-vanishing exponents occurring in the Fourier expansions of any of the n components of any solution are linear combinations, with positive integral coefficients, of a finite number of the exponents λ_m occurring in* (4).

Let the coefficient matrix, (f_{ik}), of (1), be denoted by F (and, correspondingly, the matrix of the Fourier expansions (4) by

$$F(s) \sim \sum_m A_m e^{-\lambda_m s}, \tag{5}$$

where $(a_m{}^{ik}) = A_m$). Then (1) can be written in the form

$$dW/ds = F(s)W, \tag{6}$$

if W denotes a matrix, $W = (w_{ik})$, in which each of the n columns represents a solution vector of (1).

If $W(s)$ is a fundamental matrix of (1), i. e., a matrix formed by n linearly independent solution vectors, the most general fundamental matrix of (1) is $W(s)C$, where C is any constant matrix of non-vanishing determinants. Hence, all but the last of the assertions of the Theorem will be proved if it is shown that (6) has a solution $W(s)$ which is uniformly almost-periodic in the half-plane $\sigma \geqq 0$ and is such that the Fourier expansion of

$$W(s) - E \qquad (E = \text{th} \tag{7}$$

contains no constant term. (The truth of the last assertion of the Theorem, i. e., of the assertion that every exponent of the Fourier expansion of (7) is a linear combination, with positive integral coefficients, of the exponents occurring in (5), will be clear from the proof below.)

It can be assumed that

$$\lambda_m \geqq 1 \tag{8}$$

in (5). For, on the one hand, it is assumed that $\lambda_m \geqq p$ holds for every m and for some $p > 0$, and, on the other hand, the assumptions and the assertions of the Theorem remain unaltered if s is replaced by s/p, where p is any positive constant.

3. The comments made after (3) imply that the proof of the Theorem must involve deep function-theoretical properties of the uniformly almost-periodic functions of a complex variable. The properties in question will enter via the following Lemma:

LEMMA. *On the half-plane* $\sigma \geqq 0$, *let* $f(s)$ *be a uniformly almost-periodic function, with a Fourier expansion*

$$f(s) \sim \sum_m a_m e^{-\mu_m s}, \tag{9}$$

satisfying

$$\mu_m \geqq \nu \text{ for some } \nu > 0. \tag{10}$$

Then $f(s)$ *is bounded, say*

$$|f(s)| \leqq c \text{ for } \sigma \geqq 0, \tag{11}$$

and there exists on the half-plane $\sigma \geqq 0$ *a uniformly almost-periodic function,* $f^*(s)$, *having the Fourier expansion*

$$f^*(s) \sim -\sum_m a_m \mu_m^{-1} e^{-\mu_m s} \tag{12}$$

and satisfying the inequality

$$|f^*(s)| \leqq c\nu^{-1} e^{-\nu\sigma} \text{ if } \sigma \geqq 0; \tag{13}$$

furthermore,

$$df^*(s)/ds = f(s) \text{ if } \sigma > 0. \tag{14}$$

First, the boundedness of (9) in the half-plane $\sigma \geqq 0$, i. e., the existence of a $c < \infty$ satisfying (11), follows from the assumption that $f(s)$ is almost-periodic on the closed half-plane $\sigma \geqq 0$ and from (9) and (10). In fact, this is contained in a theorem of Bohr (cf. [1], pp. 150-151). Next, (11), (10) [redacted] imply that

(15) $$|f(s)| \leqq ce^{-\nu\sigma} \text{ if } \sigma \geqq 0$$

(cf. [1], p. 152; although the wording is slightly different there, it follows from the Phragmén-Lindelöf principle that (15) is equivalent to what is deduced loc. cit.).

Since $f(s)$ is continuous for $\sigma \geqq 0$ and regular for $\sigma > 0$, it follows from (15) that it is possible to define for $\sigma \geqq 0$ a function, $f^*(s)$, which is continuous for $\sigma \geqq 0$ and regular for $\sigma > 0$, by placing

(16) $$f^*(s) = -\int_s^\infty f(\zeta)\,d\zeta.$$

If the integration path is chosen to be parallel to the real axis, then (13) follows from (15) and (16).

Since (14) is clear from (16), all that remains to be ascertained is that the function $f^*(s)$, defined by (16), is uniformly almost-periodic for $\sigma \geqq 0$, with (12) as its Fourier expansion, if (9) is uniformly almost-periodic for $\sigma \geqq 0$ and satisfies (10). But this can be concluded from the Bohl-Bohr theorem (cf. [1], p. 7 and p. 152). In fact, it follows from (15) and Cauchy's theorem, that the integration path in (16) can be chosen to be $J + L$, where J denotes the segment joining $\zeta = s$ with $\zeta = \sigma$, where $s = \sigma + it$, and L denotes the real half-line $\sigma \leqq \zeta < \infty$; cf. [1], p. 153.

4. Due to the function-theoretical facts contained in the Lemma, the proof of the Theorem can be carried out so as to become formally about the same as the proof given in [2] for the case in which (2) takes the place of (3). In fact, the proof can now be based, as in [2], on successive approximations, as follows:

Under the assumption that the matrix (5) is uniformly almost-periodic for $\sigma \geqq 0$ and satisfies (8), put

(17) $$V_0(s) \equiv E,$$

where E is the unit matrix, and let the matrices $V_1(s), V_2(s), \cdots$ be defined, save for additive constants, by the recursion formula

(18) $$dV_{m+1}(s)/ds = F(s)\,V_m(s).$$

It will be shown that the additive constants can be determined in such a way that $V_1(s), V_2(s), \cdots$ satisfy the "initial conditions"

$$(19) \qquad V_m(\infty) = 0, \text{ where } V_m(\infty) = \lim_{\sigma \to \infty} V_m(s) \qquad (m = 1, 2, \cdots)$$

(the 0 in (19) denotes the zero matrix), and that $V_m(s)$ [redacted]es uniformly almost-periodic in the half-plane $\sigma \geqq 0$.

For any matrix $C = (c_{ik})$, where $i = 1, \cdots, n$ and $k = 1, \cdots, n$, let $| C |$ denote the greatest of the n^2 non-negative numbers $| c_{ik} |$. Then

$$(20) \qquad | AB | \leqq n | A | | B |$$

holds for any pair of n-rowed matrices A, B.

Since $F(s)$ is supposed to be uniformly almost-periodic for $\sigma \geqq 0$, with a Fourier expansion (5) satisfying (8), it follows from the first assertion, (11), of the Lemma that $| F(s) | \leqq nc$ holds for $\sigma \geqq 0$ and for some $c = \text{const.}$ But (11) and (10) imply (15). Consequently, from (8),

$$(21) \qquad | F(s) | \leqq cne^{-\sigma} \text{ if } \sigma \geqq 0.$$

It follows therefore from (17) and from the Lemma that the case $m = 0$ of (18) is satisfied by the function $V_1(s) = F^*(s)$, which is uniformly almost-periodic for $\sigma \geqq 0$ (and differentiable, i. e., regular, for $\sigma > 0$); that this $V_1(s)$ satisfies the inequality

$$(22) \qquad | V_1(s) | \leqq cne^{-\sigma} \text{ if } \sigma \geqq 0;$$

finally, that all exponents occurring in the Fourier expansion of $V_1(s)$, being the same as those occurring in (5), are subject to the inequality (8).

Suppose that, for a fixed $m \geqq 1$, the function $V_m(s)$ has been defined in such a way that it is uniformly almost-periodic for $\sigma \geqq 0$, and has a Fourier expansion which does not contain any exponent less than m. It is then clear from (5) and (8) that $F(s)V_m(s)$ is uniformly almost-periodic for $\sigma \geqq 0$, with a Fourier expansion in which no exponent is less than $m + 1$. Hence, if it is assumed that the given $V_m(s)$ satisfies the inequality

$$(23) \qquad | V_m(s) | \leqq (nc)^m e^{-m\sigma}/m! \text{ if } \sigma \geqq 0,$$

then, since (23), (21) and (20) imply that

$$| F(s) V_m(s) | \leqq n(nc)^{m+1} e^{-(m+1)\sigma}/m! \text{ if } \sigma \geqq 0,$$

the Lemma supplies the existence of a $V_{m+1}(s)$ having the following properties: $V_{m+1}(s)$ is uniformly almost-periodic for $\sigma \geqq 0$, satisfies (18) if $\sigma > 0$, and

$$| V_{m+1}(s) | \leqq n(nc)^{m+1}(m+1)^{-1}e^{-(m+1)\sigma}/m! \text{ if } \sigma \geqq 0.$$

Since the inequality $| F(s) | \leqq nc$, where $\sigma \geqq 0$, was the only restriction on the choice of c, it can be assumed that $c \geqq 1$. Then the last formula line implies [redacted] (23) remains true if m is replaced by $m+1$. On the other hand, [redacted] that (23) is true for $m=1$ if $c \geqq 1$. This completes the induction (in particular (19) holds, by (23), for every $m \geqq 1$).

5. Put

$$W(s) = E + \sum_{m=0}^{\infty} \{ V_{m+1}(s) - V_m(s) \} \qquad (V_0 = E). \tag{24}$$

In view of (23), the series (24) is uniformly convergent in the half-plane $\sigma \geqq 0$. Its sum, $W(s)$, is uniformly almost-periodic for $\sigma \geqq 0$, since every term of (24) is. Furthermore, (23) and (18) show that the derived series,

$$\sum_{m=0}^{\infty} \{ V_{m+1}'(s) - V_m'(s) \},$$

is uniformly convergent for $\sigma \geqq 0$. Hence, the function (24) is differentiable on the closed half-plane $\sigma \geqq 0$ and, in view of (18), satisfies (6). Finally, since the Fourier expansion of $V_m(s)$ contains no exponent less than m, (24) shows that the Fourier expansion of the difference (7) has no constant term.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. S. Besicovitch, *Almost Periodic Functions*, Cambridge, 1932.

[2] A. Wintner, "On the classical existence theorem of linear differential equations," *American Journal of Mathematics*, vol. 71 (1949), pp. 331-338.

SEPARATION THEOREMS FOR BOUNDED HERMITIAN FORMS.*

By PHILIP HARTMAN and AUREL WINTN

---

1. If $2a_1 \leqq 2a_2 \leqq \cdots \leqq 2a_n$ are the lengths of the axes of an n-dimensional ellipsoid, and $2b_1 \leqq 2b_2 \leqq \cdots \leqq 2b_{n-1}$ those of its projection on a hyperplane through the center, then $a_1 \leqq b_1 \leqq a_2 \leqq \cdots \leqq a_{n-1} \leqq b_{n-1} \leqq a_n$. A corresponding theorem holds for the $(n-1)$-dimensional projections of a quadric belonging to a quadratic form, say $Q_n(x_1, x_2, \cdots, x_n)$, which is indefinite or semi-definite. This is just a restatement of a separation theorem of Sturm, according to which the characteristic numbers, say $\lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_n$ and $\mu_1 \leqq \mu_2 \leqq \cdots \leqq \mu_{n-1}$, of $Q_n(x_1, x_2, \cdots, x_n)$ and $Q_n(0, x_2, \cdots, x_n)$ satisfy the inequalities

$$\lambda_1 \leqq \mu_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_{n-1} \leqq \mu_{n-1} \leqq \lambda_n \tag{1}$$

(cf., e. g., [1], pp. 210-217 or [4], pp. 36-40). In fact, there is no loss of generality in assuming that the coordinate system $(x_1, \cdots, x_n)$ is so chosen as to make the equations of the n-dimensional quadric and of the hyperplane $Q_n(x_1, x_2, \cdots, x_n) = \text{const.}$ and $x_1 = 0$, respectively (even though Q_n then is not the given Q_n).

The classical proofs of (1) depend on the consideration of Sturmian chains, constructed in terms of polynomials attached to the secular determinants. This apparatus is not available if $Q_n = Q_n(x_1, \cdots, x_n)$ is replaced by a quadratic form in an infinity of variables, $Q = Q(x_1, x_2, \cdots)$; not even if the latter is assumed to be completely continuous (in Hilbert's sense) but is otherwise arbitrary. On the other hand, it is known that, *if* $Q(x_1, x_2, \cdots)$ *is completely continuous,* then its spectrum consists of 0 and of an infinite sequence of eigenvalues which tend to 0; that 0 may or may not be an eigenvalue (though it is always in the spectrum); finally, that every non-vanishing eigenvalue is of finite multiplicity (cf. [3], p. 148). In addition, Hilbert has proved that, if $Q_n = Q_n(x_1, x_2, \cdots, x_n)$ denotes the n-th section, $Q(x_1, \cdots, x_n, 0, 0, \cdots)$, of a completely continuous $Q = Q(x_1, x_2, \cdots)$, then, as $n \rightarrow \infty$, the spectrum of Q_n tends to the spectrum of Q (even if the multiplicity of the eigenvalues of Q are counted); cf. [3], pp. 156-174. Finally, it is clear that the complete continuity of $Q(x_1, x_2, \cdots)$ assures

---

* Received January 14, 1949.

that of $Q(0, x_2, x_3, \cdots)$. These facts imply that, if λ', λ'', λ^* and μ', μ'', μ^* denote points in the spectrum of $Q(x_1, x_2, \cdots)$ and $Q(0, x_2, x_3, \cdots)$, respectively, then there belong to every pair λ', λ'' satisfying $\lambda' \leqq \lambda''$ some μ^* satisfying $\lambda' \leqq \mu^* \leqq \lambda''$, and to every pair μ', μ'' satisfying $\mu' \leqq \mu''$, some λ^* satisfyi[illegible] $\leqq \lambda^* \leqq \mu''$. In fact, this extension of (1) to completely continuous forms results from (1) itself, if (1) is applied to the n-th section of $Q(x_1, x_2, \cdots)$, the m-th section of $Q(0, x_2, x_3, \cdots)$, and then $n \rightarrow \infty$, $m \rightarrow \infty$.

For a direct verification of this extension of (1), cf. a deduction given by Weyl [6], p. 167. His procedure fails if Q, instead of being completely continuous, is just bounded in Hilbert's sense. The failure is due not merely to the possibility of a continuous spectrum, since the method can fail when Q is orthogonally equivalent to a diagonal form. Actually, Weyl's method is readily seen to fail whenever the spectrum of Q has at least two cluster points (in the completely continuous case, 0 is the only such cluster point).

Although the above procedure, consisting of (1) and of a limit process, applies in certain cases in which Weyl's method fails, it does not apply to every bounded Q. The trouble is that the spectra of the sections Q_n of a bounded Q are capable of clustering at a value λ which is not in the spectrum of Q. This is shown by the example

$$2x_1x_2 + 2x_3x_4 + 2x_5x_6 + \cdots \tag{2}$$

(Toeplitz). In fact, the matrix of the bounded quadratic form (2) is an orthogonal matrix, and so the spectrum of (2), being situated both on the real axis and on the unit circle, cannot contain points distinct from $\lambda = \pm 1$ (incidentally, both of these points are eigenvalues of infinite multiplicity). On the other hand, the discriminant of the n-th section of (2) vanishes when $n = 1, 3, 5, \cdots$. Hence, $\lambda = 0$ is a cluster value of the spectra of the sections of (2), although $\lambda = 0$ is not in the spectrum of (2).

Accordingly, an extension of (1) to arbitrary bounded forms cannot be obtained from (1) itself. It will, however, be shown that such an extension exists. Needless to say, the method to be applied will have to be distinct from the determinantal approach, based on Sturmian chains, as well as from Weyl's approach, referred to above.

2. Let H be a bounded Hermitian matrix, x a point of the complex Hilbert space, and e an x of length 1. Project the Hermitian form x^*Hx on the hyperplane $x^*e = 0$ and denote by ${}^0H = {}^0H(e)$ the matrix of an Hermitian form which thus results on $x^*e = 0$. For instance, if $H = (h_{nm})$, where $n, m = 1, 2, \cdots$, and if, without loss of generality, the hyperplane

$x^*e = 0$ is chosen to be a coordinate hyperplane, $x_1 = 0$, then ${}^0H = (h_{n+1\ m+1})$, where $n, m = 1, 2, \cdots$.

The theorem announced can now be formulated as follows:

(*) *Let* ${}^0H = {}^0H(e)$ *denote any of the above-defined projections of a bounded Hermitian matrix* H. *Let* λ' *and* λ'', *where* $\lambda' <$ [illegible], *be* [illegible] *pair of points contained in the spectrum of one of the two matrices* H, ${}^0H$. *Then the spectrum of the other matrix has at least one point on the closed interval* $[\lambda', \lambda'']$.

Both of these assertions remain true in the limiting case, $\lambda' = \lambda''$, (*where the interval* $[\lambda', \lambda'']$ *becomes the point* λ') *if* λ' *is either a cluster point of the spectrum, or a multiple point of the point spectrum, of* H, ${}^0H$, *respectively.*

The (bounded, Hermitian) matrix ${}^0H = {}^0H(e)$, operating on the given linear subspace,

$$e^*x = 0, \qquad (|e| = 1), \tag{3}$$

of Hilbert's space, can explicitly be represented as

$${}^0H = (I - E)H(I - E), \tag{4}$$

where I is the unit matrix and $E = E(e)$ denotes the matrix defined by the linear substitution

$$Ex = (e^*x)e, \qquad (|e| = 1). \tag{5}$$

In fact, the latter represents the operation of projecting the x-space on the normal of the hyperplane (3).

Since every bounded linear form is completely continuous, it is seen from (5) that the matrix E is completely continuous and so, in particular, bounded. It is also seen from (5) that the matrix E is Hermitian, $E^* = E$. This was used in (4), where $H^* = H$, $I^* = I$. Finally, since repeated application of (5) gives $E[Ex] = (e^*[e^*x]e)e$, which, in view of $e^*e = 1$, is equivalent to $E[Ex] = [e^*x]e$, it follows that $E^2x = Ex$ is an identity. This merely verifies the fact that the Hermitian matrix E, representing a projection, must be idempotent, $E^2 = E$.

3. The formal basis of the proof of (*) will be the fact that

$$x^*({}^0H)^2x = x^*H^2x - |x^*He|^2 \tag{6}$$

is an identity on the hyperplane (3). This identity can be verified as follows:

If x is on the hyperplane (3), it is seen from (5) that $Ex = 0$. In view of (4), this implies that ${}^0Hx = (I - E)Hx$. It follows therefore from (4) that ${}^0H({}^0Hx) = (I - E)H(I - E)^2Hx$. But two applications of $E^2 = E$ reduce this relation to ${}^0H({}^0Hx) = (H - EH)^2x$. Since $H^* = H$ and [illegible] ows that the Hermitian form on the left of (6) is identical with [illegible] ${}^2x$. This, in turn, is identical with $x^*(H^2 - HEH)x$, since $Ex = 0$. Accordingly, (6) is equivalent to $x^*HEHx = |x^*He|^2$. But $E = E^2$ and $(EH)^* = HE$, hence $x^*HEHx = |EHx|^2$. On the other hand, the definition, (5), of E shows that $|EHx| = |e^*Hx|$. Consequently, (6) is equivalent to $|x^*He| = |e^*Hx|$..

This completes the proof of (6), since, H being Hermitian, $x^*He = (e^*Hx)^*$.

4. In the proof of (*), it can be assumed that the mid-point of the interval $\lambda' \leqq \lambda \leqq \lambda''$ (or, if $\lambda' = \lambda''$, the point $\lambda = \lambda'$) is $\lambda = 0$, i. e., that $\lambda'' = -\lambda' \geqq 0$. In fact, this normalization, being just a shift of the origin of the λ-axis, can be effected by adding to H a scalar multiple of the unit matrix; a modification which, on the subspace (3), is paralleled by the adding to ${}^0H$ of the same scalar multiple of the respective unit matrix.

It follows that, according as $[\lambda', \lambda'']$ is a point or an interval, it can be assumed that $\lambda' = 0 = \lambda''$ or $\lambda' = -1$, $\lambda'' = 1$. In fact, the latter normalization results, if $\lambda'' = -\lambda' > 0$, by a change of the unit of length on the λ-axis.

Proof of (*) *for a given* H; $\lambda' \neq \lambda''$. In this case, the assumption is that $\lambda = -1$ and $\lambda = 1$ are in the spectrum of H. The claim is that the interval $-1 \leqq \lambda \leqq 1$ cannot be free of the spectrum of ${}^0H$.

Suppose, for a moment, that the points $\lambda = \pm 1$ of the spectrum of H are in the point spectrum of H. Then there exist two perpendicular unit vectors, say x_1 and x_2, satisfying $Hx_1 = -x_1$ and $Hx_2 = x_2$. These relations clearly imply that, if x_0 is any unit vector which is a linear combination of x_1 and x_2, then $x_0^*H^2x_0 = 1$. Since the hyperplane (3) and the hyperplane spanned out by x_1 and x_2 intersect along a linear manifold (of dimensionality 1 or 2), it is possible to choose the above $x = x_0$ so as to satisfy (3). Then, since (3) implies (6),

$$(7) \qquad x_0^*({}^0H)^2x_0 = 1 - |x_0^*He|^2 \qquad (|x_0| = 1).$$

The existence of such an x_0 was obtained under the assumption that $\lambda = -1$ and $\lambda = 1$ are in the point spectrum of H. If they are just cluster values of the point spectrum of H, it is clear from the proof of (7) that the assertion of (7) becomes true if it is modified as follows: Corresponding to every $\epsilon > 0$, there exists a unit vector $x_0 = x_0^\epsilon$ satisfying

$$(8) \qquad x_0^*({}^0H)^2 x_0 < 1 - | x_0^* He |^2 + \epsilon \qquad (| x_0 | = 1, \ x_0 = x_0^\epsilon).$$

On the other hand, if $\lambda = -1$ and $\lambda = 1$ are in the con [illegible] of H, then, instead of the (weak) clustering of eigensolutions proper, an application of Hellinger's eigendifferentials (cf. [2], pp. 240-242) leads to the existence of an $x_0 = x_0^\epsilon$ satisfying (8), where $\epsilon > 0$ is arbitrary. Finally, it is clear that the same holds if $\lambda = -1$ and $\lambda = 1$ are not in one and the same of the three possible components (viz., point spectrum, cluster set of the point spectrum, continuous spectrum) of the spectrum of H.

Since $-| x_0^* He |^2 \leqq 0$, it follows, by letting $\epsilon \to 0$ in (8), that the greatest lower bound of the Hermitian form $x^*({}^0H)^2 x$ on the sphere $| x | = 1$ cannot exceed 1. On the other hand, this greatest lower bound cannot be negative, since the form is identical with $| {}^0Hx |^2$. But this greatest lower bound always is in the spectrum of the form (cf. [7], p. 147). Consequently, some point of the interval $0 \leqq \lambda \leqq 1$ must be in the spectrum of $({}^0H)^2$. In view of Toeplitz's criterion for the existence or non-existence of a bounded rceiprocal (cf. [7], p. 138), this implies that some point of the interval $-1 \leqq \lambda \leqq 1$ must be in the spectrum of ${}^0H$.

This proves the first assertion of (*).

Proof of (*) *for a given* ${}^0H$; $\lambda' \neq \lambda''$. The assumption now is that $\lambda = -1$ and $\lambda = 1$ are in the spectrum of ${}^0H$. The corresponding assertion of (*) will be proved it if is shown that, on the sphere $| x | = 1$, the greatest lower bound of the non-negative definite form $x^* H^2 x$ cannot be greater than 1. For, if this is proved, it follows, as above, that some point of the interval $0 \leqq \lambda \leqq 1$ is in the spectrum of H^2, which, in view of Toeplitz's criterion, implies that some point of the interval $-1 \leqq \lambda \leqq 1$ is in the spectrum of H.

Accordingly, it is sufficient to prove that, if $\lambda = -1$ and $\lambda = 1$ are in the spectrum of ${}^0H$, then, corresponding to every $\epsilon > 0$, some unit vector $x_0 = x_0^\epsilon$ must satisfy the inequality $x_0^* H^2 x_0 < 1 + \epsilon$. Actually, it will be sufficient to show that, if $\lambda = -1$ and $\lambda = 1$ are in the point spectrum of ${}^0H$, then some unit vector x_0 must satisfy the equality $x_0^* H^2 x_0 = 1$. For, if this is proved, it will be clear that the transition to the general case, in which both points $\lambda = \pm 1$ are in the spectrum (but not necessarily in the point spectrum) of H, requires exactly the same steps as the transition from (7) to (8) above.

Let $\lambda = -1$ and $\lambda = 1$ be in the point spectrum of $^\circ H$. Then there exists a pair of perpendicular unit vectors, say x_1 and x_2, satisfying $^\circ H x_1 = -x_1$ and $^\circ H x_2 = x_2$. Clearly,

$$x_0{}^*(^\circ H)^2 x_0 = 1 \qquad (|x_0| = 1)$$

holds [illegible] ctor, x_0, contained in the linear manifold spanned out by x_1 and x_2. Such an x_0 can be found in any given hyperplane of the x-space. Let this hyperplane be given as

$$x_0{}^* H e = 0,$$

where e is the unit vector to which $^\circ H = {}^\circ H(e)$ belongs.

Since x_1 and x_2 are eigenvectors of $^\circ H$, and since $^\circ H$, being the result of projecting H, operates only on vectors contained in the hyperplane (3), it is understood that any linear combination of x_1 and x_2, and therefore the above x_0, must satisfy (3). But (3) implies (6). Hence, (6) is satisfied by $x = x_0$. It follows therefore from the last two formula lines that $1 = x_0{}^* H^2 x^*$, where $|x_0| = 1$. This is what had to be proved (for some x_0).

Proof of (*) *in the case* $\lambda' = \lambda''$. Let K denote either of the matrices H, $^\circ H$, and L the other matrix. Then both assertions made by (*) for the limiting case, $\lambda' = \lambda''$, can be formulated as follows: The point $\lambda = 0$ must be in the spectrum of K if it is either a cluster point of the spectrum of L or a multiple eigenvalue of L. In view of what has been proved above for the case of an arbitrary closed interval $[\lambda', \lambda'']$, where $\lambda' < \lambda''$, only the second of the contingencies, that of a multiple eigenvalue, need be considered; simply because an $[\lambda', \lambda'']$, with either $\lambda' = 0$, $\lambda'' > 0$ or $\lambda' < 0$, $\lambda'' = 0$, can always be found if $\lambda = 0$ is a cluster point.

Accordingly, it is sufficient to show that $Kx = 0$ must have a solution $x = x_0$, where $|x_0| = 1$, if $Lx = 0$ has at least two linearly independent solutions, say $x = x_1$ and $x = x_2$. The latter can be assumed to be perpendicular unit vectors. But if they are used in the same way as the two eigenvectors x_1, x_2 were used in the above pair of proofs in the case $\lambda' \neq \lambda''$, it is clear that (6) leads, via (3), to a unit vector x_0 satisfying $x_0{}^*(H^\circ)^2 x_0 = 0$ or $x_0{}^* H^2 x_0 = 0$, according as $L = H$ or $L = {}^\circ H$. In other words, there exists, in either case, a unit vector x_0 satisfying $x_0{}^* K^2 x_0 = 0$. Since $x_0{}^* K^2 x_0 = |Kx_0|^2$, this proves that $Kx_0 = 0$.

5. Sturm's separation theorem, (1), can be restated as follows: If $\lambda_1{}^n \leqq \lambda_2{}^n \leqq \cdots \leqq \lambda_n{}^n$ are the roots of the secular equation of the n-th section, $H^{(n)} = \{(h_{ik}); 1 \leqq i, k \leqq n\}$, of an infinite Hermitian matrix, $H = \{(h_{ik}); 1 \leqq i, k < \infty\}$, then

$$\lambda_1{}^n \leqq \lambda_1{}^{n-1} \leqq \lambda_2{}^n \leqq \lambda_2{}^{n-1} \leqq \cdot\cdot\cdot \leqq \lambda_{n-1}{}^n \leqq \lambda_{n-1}{}^{n-1} \tag{9}$$

$(n = 2, 3, \cdots)$. Since λ_{1n} is the minimum, and $\lambda_n{}^n$ the the Hermitian form of $H^{(n)}$ on the unit sphere, it is clear from Hilbert's definition of the boundedness of H that H is bounded if and only if the values occurring in the infinite triangular matrix $\{(\lambda_m{}^n; 1 \leqq m \leqq n < \infty\}$ are contained in a bounded interval. It follows therefore from (9), and from the convergence of every bounded monotone sequence, that if $m = 1, 2, \cdots$, then

$$\lambda'_m = \lim_{n \to \infty} \lambda_m{}^n \tag{10}$$

and

$$\lambda''_m = \lim_{n \to \infty} \lambda_{n-m+1}{}^n \tag{10 bis}$$

exist. Furthermore, by (9), (10) and (10 bis),

$$\lambda'_1 \leqq \lambda'_2 \leqq \lambda'_3 \leqq \cdot\cdot\cdot \leqq \lambda''_3 \leqq \lambda''_2 \leqq \lambda''_1. \tag{11}$$

Finally, (11) shows that

$$\lim_{n \to \infty} \lambda'_m \tag{12}$$

and

$$\lim_{n \to \infty} \lambda''_m \tag{12 bis}$$

exist, and that the value of (12) does not exceed that of (12 bis).

If s is the spectrum, c the continuous spectrum, and p the point spectrum, of a bounded H, let s_0 denote the set of those points of s which are contained in one, at least, of the following three sets: c, the derivative of p, and those points of p corresponding to which H has an infinity of linearly independent characteristic vectors. In other words, the set s_0, the so-called essential spectrum of H, is the derivative of s, with the proviso that every proper eigenvalue of infinite multiplicity is counted as a cluster point of s. Clearly, the essential spectrum of H is bounded and not vacuous. It is also clear that H is completely continuous if and only if its essential spectrum consists of the single point $\lambda = 0$.

The sectional spectra $(\lambda_1{}^n, \cdots, \lambda_n{}^n)$ are not of course unitary invariants of H. It is this circumstance which makes possible a situation of the type described after (2). Situations of that type are, however, strongly limited by the following theorem:

(**) *Let* λ' *denote the least, and* λ'' *the greatest, value occurring in the esse*[illegible] *a bounded Hermitian matrix* H, *and let* $(\lambda_1{}^n, \lambda_2{}^n, \cdots, \lambda_n{}^n)$ *be t*[illegible]*e n-th section,* $H^{(n)}$, *of* H. *Then the limit* (12), *defined by* (10), *is identical with* λ'.

Furthermore, if no eigenvalue of H *is less than* λ', *then each of the limits* λ'_m *equals* λ'. *If* H *has an infinite sequence of (possibly multiple) eigenvalues which are less than* λ', *then this sequence is identical with the sequence* $\lambda'_1 \leqq \lambda'_2 \leqq \cdots$. *Finally, if* H *has a finite number, say* l, *of (not necessarily distinct) eigenvalues which are less than* λ', *then these eigenvalues are represented by the* l *values* $\lambda'_1 \leqq \cdots \leqq \lambda'_l$, *whilst* $\lambda'_{l+1}, \lambda'_{l+2}, \cdots$ *become equal to* λ'.

Corresponding statements hold for (10 bis), (12 bis) *and* λ''.

For the case of a completely continuous H, the assertions of (**) were proved by Schur [5], p. 297. He has also shown (loc. cit., p. 291) that, if H is just bounded, then (10), (10 bis), hence (12), (12 bis), are unitary invariants of H. The assertions of (**) determine the spectral meaning of these unitary invariants. Needless to say, (**) contains the unitary invariance of (10), (12).

In the proof of (**), use will be made of the set, say s^*, of the cluster values of the sectional spectra, $(\lambda_1{}^n, \cdots, \lambda_n{}^n)$. In other words, s^* will denote the set of those λ-values corresponding to which there exist two sequences, say $n_1, n_2, \cdots$ and $m_1, m_2, \cdots$, satisfying

$$(13) \qquad \lambda_m{}^n \to \lambda \text{ as } k \to \infty, \text{ where } n = n_k \to \infty; \; m = m_k.$$

Clearly, s^* is a closed set. It is known that the spectrum of H is in s^* (cf. [7], p. 218), and that the least and the greatest values of s^* are contained in the spectrum of H (ibid., p. 147). On the other hand, the example (2) shows that not every point of s^* need be in the spectrum of H.

6. Consider first the case in which, on the one hand, the least value, λ', contained in the essential spectrum, s_0, is the least value contained in the spectrum, s, and, on the other hand, λ' is not in the point spectrum. The first of these assumptions means that $l = 0$; cf. the wording of (**). In view of the second asumption, this implies that the interval $\lambda' < \lambda < \lambda' + \epsilon$ contains points of s whenever $\epsilon > 0$. Furthermore, λ' is the least value contained in s^*. Hence it is clear from (10) that $\lambda'_1 = \lambda'$.

It will be shown that $\lambda'_m = \lambda'$ holds for every m. In view of (11), this will prove the assertions of (**) under the present pair of assumptions.

Suppose, if possible, that $\lambda' \neq \lambda_m$, hence $\lambda'_1 < \lambda'_m$, holds for some m. Then s^* has at most $m - 2$ points, hence just a finite nu ts, on the interval $\lambda'_1 < \lambda < \lambda'_m$. Consequently, this interval in an infinity of points of s. Hence, $\lambda' = \lambda'_1$ is not a cluster po ace it is assumed that λ' is not in p, it follows that λ' is not in s_0. But this contradicts the definition of λ'.

7. Without the particular assumptions made at the beginning of § 6, the proof of (**) proceeds as follows:

Let $\lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda'$ denote the finite or infinite (possibly vacuous) sequence of those eigenvalues of H which are not greater than λ', the least value contained in s_0. Then λ_1 is the least value contained in s^*, and so $\lambda_1 = \lambda'_1$.

Suppose, therefore, that $\lambda_1 = \lambda'_1, \cdots, \lambda_m = \lambda'_m$ has been proved for a fixed m. Suppose further that, in the sequence $\lambda_1, \lambda_2, \cdots$, there occurs an $(m+1)$-st eigenvalue, λ_{m+1}. It will be shown that

$$\lambda_{m+1} = \lambda'_{m+1}. \tag{14}$$

If $u = (u_1, u_2, \cdots)$ is a point in Hilbert's space, let ${}^n u$ denote either the point $(u_1, u_2, \cdots, u_n, 0, 0, \cdots)$ of the same space or the point $(u_1, u_2, \cdots, u_n)$ of the n-dimensional space. This two-fold meaning of ${}^n u$ will not lead to a confusion below. For the sake of convenience, let linear forms, written above as matrix products, be now written as scalar products, (u, v).

8. Under the assumptions made before (14), let $e_1, e_2, \cdots, e_m, e_{m+1}$ denote mutually perpendicular unit vectors which are eigenvectors belonging to $\lambda_1, \lambda_2, \cdots, \lambda_m, \lambda_{m+1}$, respectively. For every fixed $n > m$, consider the problem of minimizing the n-th section, ${}^n x^* H {}^n x = (H\, {}^n x, {}^n x)$, of $x^* H x = (Hx, x)$ under the $1 + m$ conditions

$$|\,{}^n x\,| = 1 \text{ and } ({}^n e_j, {}^n x) = 0, \text{ where } j = 1, \cdots, m; \qquad n > m.$$

According to the so-called Rayleigh principle, which simply follows by a repeated application of (9), this minimum is not greater than $\lambda_{m+1}{}^n$. On the other hand, since $(H\, {}^n x, {}^n x) = (Hx, x)$ and $(e_j, {}^n x) = 0$, where $j = 1, \cdots, n$, the definition of λ_{m+1} shows that the minimum in question is not less than λ_{m+1}. Consequently, $\lambda_{m+1} \leqq \lambda_{n+1}{}^n$.

Since the last inequality and (10) imply that $\lambda_{m+1} \leqq \lambda'_{m+1}$, the proof of (14) will be complete if it is shown that $\lambda_{m+1} \geqq \lambda'_{m+1}$.

9. To this end, let $f_1^n, \cdots, f_n^n$ be a (complete) orthonormal set of eigenvectors of the n-th section, $H^{(n)}$, of H (i. e., let

$$H^{(n)} f_i^n = \lambda_i^n f_i^n, \qquad i = 1, \cdots, n; \qquad (f_i^n, f_h^n) = (\epsilon_{ih}),$$

wher[illegible] n-rowed unit matrix), and let n be greater than the m occurring in the given λ_{m+1}. If $e_1, \cdots, e_{m+1}$ are defined as at the beginning of § 8, choose $m+1$ scalars $\alpha_1^n, \cdots, \alpha_{m+1}^n$ in such a way that the vector

$$g_n = \sum_{k=1}^{m+1} \alpha_k^n e_k$$

satisfies the $1 + m$ conditions

$$|\, g_n \,| = 1, \qquad (g_n, f_i^n) = 0, \qquad\qquad i = 1, \cdots, m.$$

It is readily seen from this definition of the points $x = g_{m+1}, g_{m+2}, \cdots$ of Hilbert's space that they must strongly cluster at some point, say at $x = d$ (in fact, only a finite number, $m+1$, of dimensions need be considered). In other words, if a suitable subsequence of $g_{m+1}, g_{m+2}, \cdots$ is denoted simply by $g_{m+1}, g_{m+2}, \cdots$, then $|\, g_n - d \,| \to 0$ as $n \to \infty$. On the other hand, by the definition of the frontal superscript, $|\, {}^n d - d \,| \to 0$. Since

$$({}^n d, f_i^n) = ({}^n d - d, f_i^n) + (d - g_n, f_i^n) + (\text{[illegible]}, f_i^n),$$

it follows that

$$({}^n d, f_i^n) \to 0 \text{ as } n \to \infty, \text{ where } i = 1, \cdots, m.$$

Since λ_{m+1}^n is the $(m+1)$-st eigenvalue of $H^{(n)}$, and since $(H\, {}^n x, {}^n x) \to (Hx, x)$ as $n \to \infty$ whenever $|\, x \,| < \infty$, it follows that

$$(Hd, d) \geqq \liminf_{m \to \infty} (\lambda_{m+1}^n \,|\, {}^n d \,|^2).$$

But $\lim |\, {}^n d \,| = |\, d \,|$. Furthermore, $|\, d \,| = 1$, since $|\, g_n \,| = 1$ and $\lim |\, g_n - d \,| = 0$. It follows therefore from (10), and from the last formula line, that $(Hd, d) \geqq \lambda'_{m+1}$.

Finally, it is clear from the definitions of λ_{m+1}, g_n and d that $(Hd, d) \leqq \lambda_{m+1}$. It follows therefore from the preceding inequality that $\lambda'_{m+1} \leqq \lambda_{m+1}$. Since $\lambda_{m+1} \leqq \lambda'_{m+1}$ was proved in § 8, the proof of (14) is now complete.

Consequently, if the sequence $\lambda_1 \leqq \lambda_2 \leqq \cdots (\leqq \lambda')$ of eigenvalues not exceeding λ' is infinite, the assertions (**) follow. On the other hand, if this sequence contains only a finite number, l, of eigenvalues, the proof of (**) can be concluded by an application of the arguments used in § 6.

APPENDIX.

t s be the spectrum of a bounded Hermitia
$_1{}^n, \cdots, \lambda_n{}^n$) the spectrum of the n-th section, H(
defined near the end of § 5. The fact that s is c
gthened as follows:

λ is in s, then there exists in every s_n a point $\lambda_m{}^n$, where $m = m_n$,
atisfies

o, where $1 \leqq m = m_n \leqq n$;

it formulation is as follows:

0 and a sequence of integers $n_1 < n_2 < \cdots$
al $(\lambda - \epsilon, \lambda + \epsilon)$ contains no points of s_n,
, then λ is not in s.

d out in § 5, it is not true in general that,
ntegers $m_1, m_2, \cdots$ and $n_1, n_2, \cdots$, where
$\cdot$, and a corresponding point $\lambda_m{}^n$ of s_n
n s. This is illustrated by the H of the
s case, $\lambda_m{}^n = 0$ when $m = m_k = k + 1$ and
$= 2k + 1$, but $\lambda = \lim \lambda_m{}^n = 0$ is not in the spectrum of (2).
e idea leading to the example (2) can be refined as follows:

) *In the case of a bounded Hermitian matrix, a point λ need not be*
en if it is in every s_n; in particular, it is possible that λ is not in s
there exists in every s_n a $\lambda_m{}^n$ satisfying (15).

is possibility can be further refined:

bis) *In the case of a bounded Hermitian matrix, a point λ need*
in s even if there exist $k = k_n$ points $\lambda_{h+1}{}^n < \lambda_{h+2}{}^n < \cdots < \lambda_{h+k}{}^n$
hich satisfy

$$\lambda_{h+1}{}^n \rightarrow \lambda \text{ and } \lambda_{h+k}{}^n \rightarrow \lambda \text{ as } n \rightarrow \infty,$$

$0 \leqq h \leqq h_n \leqq n - k$ and $k = k_n \rightarrow \infty$.

ere arises the question whether the possibility admitted in (i), (i bis)
ruled out for certain types of bounded Hermitian matrices. It is
that these possibilities cannot occur for diagonal matrices. It is
to inquire whether or not they can occur in the case of Jacobi

matrices, which, according to Hellinger and Toeplitz, are the normal under unitary transformations, of matrices with simple spectra. matr the conditions

$$m; \; h_{nn+1} \neq 0; \qquad h_{nm} = 0 \text{ if } |n-m| > 1 \tag{18}$$

wher Actually, the matrix of (2), which is not of type, can be adjusted to the case of a Jacobi matrix.

(ii) *In the case of a bounded Jacobi matrix, a point* λ *need in* s *if there exists a subsequence of the sets* $s_1, s_2, \cdots$ *such that th element of the subsequence contains a point* $\lambda_m{}^n$ *which satisfies* (13), $m = m_k$ *and* $n = n_k$.

On the other hand, the possibility claimed by (i the case of a Jacobi matrix. More than this exclusic following theorem:

(iii) *In the case of a bounded Jacobi matrix, a if every* s_n *contains two points* $\lambda_m{}^n, \lambda_{m+1}{}^n$, *where* $m =$ *as* $n \to \infty$.

What is more, the interval $\lambda_m{}^n \leqq \lambda \leqq \lambda_{m+1}{}^n$ *contai of* s *for any* n *and for any pair of successive points* $\lambda_m{}^n, \lambda_{m+1}{}^n$ *of* s_n

It remains undecided whether or not the possibility in the secon of (i) can occur for a bounded Jacobi matrix; that is, whether or n remains true if "subsequence of the sets $s_1, s_2, \cdots$" is replaced by "se of the sets $s_1, s_2, \cdots$."

Proof of (i). Let the sequence of integers $1, 2, \cdots$ be divid two mutually exclusive subsequences, $i_1 < i_2 < \cdots$ and $j_1 < j_2 < \cdots$ H be the matrix defined by placing h_{mn} equal to 0 or 1 according as t (m, n) of integers does not or does have either of the forms (i_k, j_k), for some $k = 1, 2, \cdots$. It is clear that H is bounded and Hermiti that each row (column) of H contains one and only one element from 0, while the non-vanishing element is always 1. Consequently, orthogonal matrix, and so its spectrum consists of the eigenvalues λ (of infinite multiplicity). On the other hand, if i_k is chosen to be $k = 1, 2, \cdots$, then $H^{(n)}$ has at most $2[n/3]$ elements different f Thus $\lambda = 0$ is in s_n, with a multiplicity not less than $n - 2[n/3]$ It follows that the possibilities mentioned in (i) are realized by H.

roof of (i bis). This H realizes, with $k = k_n = n - 2[n/3]$, the
ility mentioned in (i bis) if the assertion (i bis) n a
sense, which allows $\lambda_{h+1}{}^n = \cdots = \lambda_{h+k}{}^n = 0$ when vith
iplicity k. It is clear, however, that there exist rea ices
h that, when H is the matrix constructed above, the
ement of (i bis). In fact, it is seen from Toe for
ded reciprocal (cf. [7], p. 238) that, if ϵ_n is the n-th diagonal element
and $\frac{1}{2} > | \epsilon_n | \rightarrow 0$, as $n \rightarrow \infty$, then $\lambda = 0$ is not in the spectrum of
). On the other hand, if n is fixed and h is an index of one of the
st) $n - 2[n/3]$ columns of $H^{(n)}$ containing only zeros, then $\lambda = \epsilon_n$
igenvalue of the n-th section of $H + D$.

natrix of the form (2) fails to be a Jacobi matrix
$= 0$ for $n = 1, 2, \cdots$. But let this matrix be
0, $h_{mn} = 0$ for $| m - n | > 1$, and $h_{2n-1,2n} = h_{2n,2n-1}$
lly $h_{2n,2n+1} = h_{2n+1,2n} = \epsilon_n$ for $n = 1, 2, \cdots$, where
The resulting matrix is a bounded Jacobi matrix.
mbers ϵ_n are suitably chosen, then the possibility
l.

clear that it is sufficient to prove the last part of
ne, the replacement of H by $H - \lambda I$ merely trans-
$- \lambda$. If λ is chosen to be $\frac{1}{2}(\lambda_m{}^n + \lambda_{m+1}{}^n)$, the
$-\lambda^*, \lambda^*$, where $\lambda^* = \frac{1}{2}(\lambda_{m+1}{}^n - \lambda_m{}^n) > 0$. Conse-
show that if $-\lambda^*$ and $\lambda^* (> 0)$ are in s_n (for a
at least one number λ satisfying $| \lambda | \leqq \lambda^*$ or,
ists a vector $x = (x_1, x_2, \cdots)$ such that

$$| Hx | \leqq \lambda^* \qquad (| x | = 1).$$

represents either a point of the type $(x_1, \cdots, x_n, 0, 0, \cdots)$ in the
t space or the corresponding point $(x_1, \cdots, x_n)$ of the n-dimensional
then, since H is a Jacobi matrix, it follows from (18) that $Hx = y$
$, y_2, \cdots)$, where

$$y_1, \cdots, y_n) = H^{(n)}x \text{ and } y_{n+1} = h_{n+1,n}x_n ; y_{n+2} = y_{n+3} = \cdots = 0.$$

quently, $| Hx |^2 = | H^{(n)}x |^2 + | h_{n+1,n}x_n |^2$. Hence,

$$| Hx | = | H^{(n)}x | \text{ if } x_n = 0.$$

n the other hand, the finite matrix $H^{(n)}$ has two eigenvalues $-\lambda^*$,

λ^* and corresponding eigenvectors x^1, x^2 satisfying $H^{(n)}x^1 =$ $H^{(n)}x$ nce, the orthogonality of x^1, x^2 implies

$$| H^{(n)}x | = \lambda^* | x |,$$

if x is nation of x^1 and x^2. Since x^1, x^2 are linearly inde the pı holds for such an x (with $| x | = 1$). The last line a (19). This completes the proof of (iii).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] R. Fricke, *Lehrbuch der Algebra*, vol. 1, Braunschweig,

[2] E. Hellinger, "Neue Begründung der Theorie quadratisch lichvielen Veränderlichen," *Journal für die reine u matik*, vol. 136 (1909), pp. 219-271.

[3] D. Hilbert, *Grundzüge einer allgemeinen Theorie der* Leipzig and Berlin, 1912.

[4] E. J. Routh, *The advanced part of a treatise on the bodies*, London, 1884.

[5] I. Schur, "Ein Beitrag zur Hilbertschen Theorie Formen," *Mathematische Zeitschrift*, vol. 12 (1922),

[6] H. Weyl, "Ueber das Spektrum der Hohlraumstrahlung," *und angewandte Mathematik*, vol. 141 (1912), pp. 163-181.

[7] A. Wintner, *Spektraltheorie der unendlichen Matrizen*, Leipzig, 1929.

AFFINE INVARIANTS OF A PAIR OF HYPERSURFACES.*

By Chuan-Chih Hsiung.

1. **Introduction.** In previous papers [1; 2; 3, p thor
has derived a projective invariant, together with ctive
characterizations, for two surfaces having a common tangent plane at two
nonsingular points in ordinary space, and has extended the result to two
hy yperspace. Santaló [4, pp. 564-569] has obtained
aff ith metric and affine characterizations, for simi-
lar es in ordinary space. The purpose of this note
is Santaló for two hypersurfaces V_{n-1}, V^*_{n-1} in n-
di having a common tangent hyperplane at two
no We shall confine our attention to the case in
wh asymptotic tangent of either of the hypersurfaces
V_n

riants. Let V_{n-1}, V^*_{n-1} be two hypersurfaces in
a nsions having a common tangent hyperplane t_{n-1}
at 0* the line 00* not being an asymptotic tangent
of es V_{n-1}, V^*_{n-1}. First, we choose an orthogonal
cartesian coordinate system with the point 0 as the origin, the line 00* as the x_1-axis, and the common tangent hyperplane t_{n-1} as the coordinate hyperplane $x_n = 0$. Then the power series expansions of the hypersurfaces V_{n-1}, V^*_{n-1} in the neighborhoods of the points 0, 0* may be written in the form

$$(1) \qquad V_{n-1}\colon\ x_n = \sum_{i,k=1}^{n-1} l_{ik} x_i x_k + \cdots,$$

$$(2) \qquad V^*_{n-1}\colon\ x_n = m_{11}(x_1 - h)^2 + 2\sum_{i=2}^{n-1} m_{1i} x_i (x_1 - h) + \sum_{i,k=2}^{n-1} m_{ik} x_i x_k + \cdots,$$

where h is the distance between the points 0, 0* and $l_{11} m_{11} \neq 0$.

In order to find the affine invariants determined by the neighborhoods of the second order of the hypersurfaces V_{n-1}, V^*_{n-1} at the points 0, 0*, we consider the most general affine transformation which leaves the point 0, the x_1-axis and the hyperplane $x_n = 0$ unchanged:

$$(3) \qquad x_i = \sum_{k=1}^{n} a_{ik} x'_k, \qquad x_n = a_{nn} x'_n \qquad (i = 1, \cdots, n-1),$$

---

* Received February 28, 1949.

where a_{21} $\cdots = a_{n-1,1} = 0$, the remaining coefficients being arbitrary, but

$$(4) \qquad = \begin{vmatrix} a_{22} & a_{23} & \cdots & a_{2,n-1} \\ a_{32} & a_{33} & \cdots & a_{3,n-1} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n-1,2} & a_{n-1,3} & \cdots & a_{n-1,n-1} \end{vmatrix} \neq 0.$$

The effect of this transformation on equations (1 two other equations of the same form whose coefficie ents, are given by the formulas

$$(5) \quad \begin{array}{l} a_{nn}l'_{11} = a^2{}_{11}l_{11}, \qquad a_{nn}l'_{ik} = \sum_{r,s=1}^{n-1} a_{rs}a_{sk}l_{rs} \qquad), \\ a_{11}h' = h, \qquad a_{nn}m'_{11} = a^2{}_{11}m_{11}, \qquad a_{nn}m'_{ik} = \qquad), \end{array}$$

where $m_{i1} = m_{1i}$ $(i = 2, \cdots, n-1)$.

From equations (4), (5) it is easily seen the

$$L = \begin{vmatrix} l_{11} & l_{12} & \cdots & l_{1,n-1} \\ l_{21} & l_{22} & \cdots & l_{2,n-1} \\ \cdot & \cdot & \cdot & \cdot \\ l_{n-1,1} & l_{n-1,2} & \cdots & l_{n-1,n-1} \end{vmatrix}, \qquad M = \begin{vmatrix} m & & {}_{1,n-1} \\ m & & {}_{2,n-1} \\ \cdot & & \cdot \\ m & & {}_{n-1,n-1} \end{vmatrix}$$

and their transformed ones L', M' are connected by the relations

$$(6) \qquad a^{n-1}{}_{nn}L' = a^2{}_{11}D^2L, \qquad a^{n-1}{}_{nn}M' = a^2{}_{11}D^2M.$$

Further elimination of a_{ik} from equations (5), (6) shows that *the two quantities*

$$(7) \qquad I = l_{11}/m_{11}, \qquad J = L/M$$

are two affine invariants determined by the neighborhoods of the second order of the hypersurfaces V_{n-1}, $V^*{}_{n-1}$ at the points 0, 0*.

3. Metric and affine characterizations of the invariants. Let K, K^* be the curvatures of the hypersurfaces V_{n-1}, $V^*{}_{n-1}$ at the points 0, 0*; and let R, R^* be the curvatures at the points 0, 0* of the plane curves C, C^* of section of the hypersurfaces V_{n-1}, $V^*{}_{n-1}$ by the plane n of the line 00* and the normal to the common tangent hyperplane t_{n-1} at any point on the line 00*. Then from equations (1), (2) and a well known formula [Cf. 5, pp. 313-314] for the curvature of a hypersurface at a nonsingular point, it is easy to obtain *the following metric characterizations for the two invariants I, J*:

$$I = R/R^*, \qquad J = K/K^*. \tag{8}$$

In order to find an affine characterization of the et us
consider any affine transformation, which transforms the V_{n-1},
V^*_{n-1} into two other hypersurfaces $\bar{V}_{n-1}$, $\bar{V}^*_{n-1}$ having th yper-
plane $\bar{t}_{n-1}$ of t_{n-1} as a common tangent hyperplane at th points
$\bar{0}$, $\bar{0}^*$ of the points 0, 0^*, and which transforms the com ane n
through the line 00^* of the hypersurfaces V_{n-1}, V^*_{n-1} into another plane m
which passes through the line $\bar{0}\bar{0}^*$ and makes an angle θ with the common
normal plane $\bar{n}$ through the line $\bar{0}\bar{0}^*$ of the hypersurfaces $\bar{V}_{n-1}$, $\bar{V}^*_{n-1}$. If
$\bar{R}$, $\bar{R}^*$ points $\bar{0}$, $\bar{0}^*$ of the plane curves $\bar{C}$, $\bar{C}^*$ of
sectio $_{-1}$, $\bar{V}^*_{n-1}$ by the plane m, then $R/R^* = \bar{R}/\bar{R}^*$,
Santa g shown that the ratio of the curvatures at
two r of two plane curves with $0_1 0_2$ as a common
tange Furthermore, if r, r^* be the curvatures at
the p curves of section of the hypersurfaces $\bar{V}_{n-1}$,
$\bar{V}^*_{n-1}$ plane $\bar{n}$, then by the well known theorem of
Meus nal space determined by the two planes m, $\bar{n}$,
we o $\cos\theta$. Thus $R/R^* = r/r^*$ and making use
of a 560-561] we therefore arrive at the following
affine variant I.

I ane n, let f be the area bounded by a line l
parall by the curve C in the neighborhood of the
point bounded by the line l (or its symmetric line
with respect to the line 00^*) *and the curve* C^* *in the neighborhood of the point* 0^*, *then* $I = \lim (f^*/f)^2$ *as the line l approaches the line* 00^*.

To characterize affinely the other invariant J we consider the pencil of hyperquadrics in the hyperplane $x_n = 0$ determined by the two asymptotic hypercones of the hypersurfaces V_{n-1}, V^*_{n-1} at the points 0, 0^*. It is easily seen that in this pencil there exist $n - 1$ hyperparaboloids which are given by the equations

$$x_n = \sum_{i,k=1}^{n-1} (l_{ik} + \lambda_j m_{ik}) x_i x_k - 2\lambda_j h \sum_{i=1}^{n-1} m_{1i} x_i + \lambda_j m_{11} h^2 = 0, \qquad (j = 1, \cdots, n-1), \tag{9}$$

where λ_j $(j = 1, \cdots, n-1)$ are the roots of the equation in λ

$$\begin{vmatrix} l_{11} + \lambda m_{11} & l_{12} + \lambda m_{12} & \cdots & l_{1,n-1} + \lambda m_{1,n-1} \\ l_{21} + \lambda m_{21} & l_{22} + \lambda m_{22} & \cdots & l_{2,n-1} + \lambda m_{2,n-1} \\ \cdots & \cdots & \cdots & \cdots \\ l_{n-1,1} + \lambda m_{n-1,1} & l_{n-1,2} + \lambda m_{n-1,2} & \cdots & l_{n-1,n-1} + \lambda m_{n-1,n-1} \end{vmatrix} = 0. \tag{10}$$

The line 00* intersects each of the hyperparaboloids (9) in a pair of points. Let P_j be each pair of these points, then

$$0^* = \pm [-(m_{11}/l_{11})\lambda]^{\frac{1}{2}} \qquad (j = 1, \cdots, n-1). \tag{11}$$

Thus *J can be expressed in terms of the invariant I and the* $n-1$ $\cdots, D_{n-1}$ *of distances between points as follows:*

$$J = (-1)^n I^{n-1} (D_1 D_2 \cdots D_{n-1})^2 \tag{12}$$

4. Discussion. Finally, it should be noted th iants I, J are not projective invariants, but the affine in is a projective invariant [3, p. 578].

Moreover, if the line 00* is an asymptotic th of the hypersurfaces V_{n-1}, V^*_{n-1}, then either $l_{11} = 0$ $= 0$. In each of these special cases we can only det nt J given by the second of equations (7) with one of the n l_{11}, m_{11}. The metric characterization of this invariant J l can be used, but the affine one no longer has a meaning.

THE UNIVERSITY OF WISCONSIN.

REFERENCES.

1. C. C. Hsiung, "Projective invariants of a pair of surfaces," *Duke Mathematical Journal*, vol. 10 (1943), pp. 717-720.
2. ———, "A projective invariant of a certain pair of surfaces," *Duke Mathematical Journal*, vol. 12 (1945), pp. 441-443.
3. ———, "Some invariants of certain pairs of hypersurfaces," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 572-582.
4. L. A. Santaló, "Affine invariants of certain pairs of curves and surfaces," *Duke Mathematical Journal*, vol. 14 (1947), pp. 559-574.
5. A. Terracini, "Densità di una corrispondenza di tipo dualistico, ed estensione dell'invariante di Mehmke-Segre," *Atti della Reale Accademia delle Scienze di Torino*, vol. 71 (1936), pp. 310-328.

COMPLETELY SIMPLE IDEALS OF A SEM

By R. P. Rich.

This note is concerned with the following question: a semigroup S contain a completely simple ideal M—that is, an ideal which is itself a completely simple semigroup?

on in terms of minimal ideals was given by
A.] ıs no nilpotent ideals and the minimal two-
side(ıinimal left ideal L and a minimal right ideal
R, t ple semigroup.
serve (Theorem A below) that the condition
that is unnecessarily restrictive and that we need
only ımption that M itself be non-nilpotent.
ıse complementary, answer to the question
(Th f S contains a minimal left ideal L and a
min ıat $LR \neq (0)$ and $RL \neq (0)$, then $M = LR$
is a ıp (and in particular a non-nilpotent minimal
two- noticed that in (A) we start with a minimal
two- itions that it be completely simple, while in
(B) and right ideals and produce a minimal two-sided ideal which is completely simple.

THEOREM A. *If M is a minimal two-sided ideal of the semigroup S, if M contains a minimal left ideal L and a minimal right ideal R, and if $M^2 \neq (0)$, then M is completely simple.*

The only ideals whose non-nilpotence is relevant to the proof given by Clifford[1] are in fact contained in M, so to establish Theorem A we need only prove

LEMMA 1. *If M is a minimal ideal of the semigroup S (with zero) and $M^2 \neq (0)$, then M contains no proper nilpotent left or right ideal of S.*

Proof. Suppose L is a proper nilpotent left ideal of S contained in M.

---

* Received February 14, 1949.

[1] A. H. Clifford, "Semigroups without nilpotent ideals," *American Journal of Mathematics*, vol. 71 (1949), pp. 834-844, Theorem 3.1. The reader is also referred to this paper for the definition of terms used here. Here, as there, S is an arbitrary semigroup with zero.

Then $L \cup LS$ is a nilpotent two-sided ideal of S contained in M and $\neq (0)$, hence [illegible] since M is minimal, and M is nilpotent, contrary to hypoth[illegible]

L[illegible] and R *are minimal left and right ideals, respectively, of the* [illegible] *(with zero) and* $LR \neq (0)$, *then:*

a) [illegible] and $0 \neq r \varepsilon R$ then $lr \neq 0$;

b) LR *is a minimal two-sided ideal of* S.

Proof. Evidently LR is a two-sided ideal of S. Let $0 \neq l' \varepsilon L$ and $0 \neq r' \varepsilon R$, with $l'r' \varepsilon T$, an ideal of S contained [illegible] of
elements r of R such that $l'r \varepsilon T$. For any r in [illegible] ve
$l'rs \varepsilon T$ since T is an ideal, so $rs \varepsilon R'$ and R' is a [illegible] n-
tained in the minimal right ideal R; hence $R' =$ [illegible] al
argument, $Lr \subseteq T$ for every $r \neq 0$ in R, so $LR \subseteq$ [illegible] In
particular, if $l'r' = 0$ let $T = (0)$ and then the [illegible] at
$LR = (0)$, contrary to hypothesis.

LEMMA 3. *If* L *and* R *are minimal left a*[illegible] *y,*
of S, *and* $LR \neq (0)$, $RL \neq (0)$, *then* $LRL = L$ [illegible] .

Proof. Let $0 \neq l_1 \varepsilon L$ and $0 \neq r_2 l_2 \varepsilon$ [illegible] ve
$l_1 \cdot r_2 l_2 \neq 0$ by Lemma 2a. Hence $LRL \neq (0)$. [illegible] S
contained in the minimal left ideal L, so $LRL = L$.

THEOREM B. *If* L *and* R *are minimal left and right ideals, respectively, of the semigroup* S *(with zero) and* $LR \neq (0)$, $RL \neq (0)$, *then* LR *is a completely simple semigroup.*

Proof. Let $M = LR$. Then M is a minimal two-sided ideal of S by Lemma 2b. By Lemma 3, $M^2 = LRL \cdot R = LR \neq (0)$. Since $RL \subseteq R$ we have $L = LRL \subseteq LR = M$, and, dually, $R \subseteq M$. Thus Theorem A applies and $M = LR$ is completely simple.

A converse of Theorem B may be stated:

THEOREM C. *If the semigroup* S *(with zero) contains a completely simple ideal* M, *then* $M = LR$, *where* L *and* R *are minimal left and right ideals, respectively, of* S *and* $RL \neq (0)$.

Proof. Let M be a completely simple ideal of S and let e be one of its primitive idempotents. Then $L = Me$ is a minimal[2] left ideal of M and a left ideal of S, hence a minimal left ideal of S. Dually, $R = eM$ is a minimal right ideal of S. Since M is simple, $LR = Me^2M =$ ly, RL contains $e^4 \neq 0$.

---

[2] *Ibid.*, proof of Theorem 3.2.

The reasoning of Theorem C shows that both $\neq (0)$, $RL \neq (0)$ are necessary conditions for Theorem B. T amples show that they are independent, and hence must both sumed.

Example 1. Let $S = (0, l, r, b)$ with $lr = b$ and all other products $= 0$. Let $L = (0, l)$, $R = (0, r)$. Then LR contains $b \neq 0$ but $RL = (0)$.

Ea $, r, s)$, with the table:

a	l	r	s
0	0	0	0
0	a	0	r
0	l	0	s
a	a	r	r
l	l	s	s.

Let L Then RL contains $a \neq 0$, but $LR = (0)$.
J imal left, R a minimal right ideal.

T Y.

ON ASSICAL EXISTENCE THEOREM OF ANALYTIC DIFFERENTIAL EQUATIONS.*

EARL A. CODDINGTON and AUREL WINTNER.

1. Let $f_1, \cdots, f_n$, where $f_i = f_i(s; w_1, \cdots, w_n)$, be n power series in $w_1, \cdots, w_n$, say

$$(1) \qquad f_i = \sum_{k=0}^{\infty} \cdots \sum_{m=0}^{\infty} a_i^k$$

where the coefficients $a = a(s)$ are

$$(2) \qquad a_i^{k \cdots m}(s) = \int_1$$

of (complex-valued) functions $\alpha(x)$ -line $1 \leqq x < \infty$. Suppose that the total

$$(3) \qquad [\alpha] = \int_1$$

of the functions $\alpha(x)$ satisfy the n conditions

$$(4) \qquad \sum_{k=0}^{\infty} \cdots \sum_{m=0}^{\infty} [\alpha_i^{k \cdots m}] r^{k + \cdots + m} < \infty \qquad (i = 1, \cdots, n)$$

for some, sufficiently small, positive r. In particular, the functions (1) are regular in the domain

$$(5) \qquad \Re s > 0, \qquad |w_1| < r, \cdots, |w_n| < r$$

of the $n + 1$ complex variables s, w_i.

The following theorem will be proved:

In a domain (5), *let* $f_1, \cdots, f_n$ *be regular functions representable, in terms of functions* (2) *satisfying* (4), *in the form* (1). *Then there exists a half-plane* $\Re s > l$ $(\geqq 0)$ *on which the system*

$$(6) \qquad dw_i/ds = f_i(s; w_1, \cdots, w_n) \qquad (i = 1, \cdots, n)$$

---

* Received November 8, 1948.

has a unique solution representable in the form

$$(7) \qquad w_i(s) = \int_1^{\infty} e^{-sx} d\beta_i(x)$$

In addition, the n integrals (7) are absolutely conv
$\Re s > l$, if l is large enough.

In particular, the solution (7) of (6) belongs to the "initial condition"

$$(8) \qquad w_1(\infty) = 0, \cdots, w_n(\infty) = 0, \text{ where } w_i(\infty) = \lim w_i(s) \text{ as } \Re s \to \infty.$$

$$\cdots \sum_{m=0}^{\infty} a_i^{jk\cdots m} z^j w_1^k \cdots w_n^m,$$

wh l existence theorem of analytic systems
of be formulated as follows: If each of
th formula line is (absolutely) convergent
in $< r, \cdots, | w_n | < r$, then there exists
a $dw_i/dz = F_i$ has a (unique, regular)
sol initial conditions $w_i(0) = 0$. This
the particular case of the theorem to be
pr icient to put $z = e^{-s}$ in (1), (2), (6),
and assume that each of the coefficient functions, $a(s)$, in (1) is a power series in e^{-s}, i. e., that each of the functions $\alpha(x)$ occurring in (2) is a step function the jumps of which are restricted to $x = 1, 2, \cdots$.

For the case in which (6) is a linear system, the above theorem was recently proved,[1] along with a refinement generalizing the fact that, in case of linear systems, the existence theorem need not be "localized" by the choice of a "sufficiently large" l. The treatment of the non-linear case will depend on an adaptation of the procedure used loc. cit.

2. In order to simplify the formulae, the proof will be given for the case, $n = 1$, of a single differential equation. It will be clear that the case of an arbitrary n can be treated in just the same manner. As a matter of fact, $n > 1$ can always be reduced to $n = 1$, if the process involved is based on the method of majorants, which will be the case below.

If $n = 1$, then (6) reduces, by (1) and (2), to the differential equation

---

[1] A. Wintner, "On the classical existence theorem of linear differential equations," *American Journal of Mathematics*, vol. 71 (1949), pp. 331-338.

$$dw/ds = \sum_{m=0}^{\infty} w^m \int_1^{\infty} e^{-sx} d\alpha_m(x), \tag{9}$$

whe 4), there exists an $r > 0$ satisfying

$$\sum_{m=0}^{\infty} [\alpha_m] r^m < \infty. \tag{10}$$

The assertion is that (10) implies for (9) the existence of a unique solution representable, on some half-plane $\Re s > l$, as an absolutely convergent Laplace-Stieltjes integral,

$$w(s) = \int_l^{\infty}. \tag{11}$$

It can of course be assumed that in (9), are normalized by

$$\alpha(1) \tag{12}$$

and

$$\alpha(x) = \alpha(x - 0) \tag{13}$$

All the functions β, β_k, to be defined below, will be understood to have been normalized by the cases $\alpha = \beta$, β_k of (12) and (13).

If (11) is substituted into (9), a "comparison of the coefficients" supplies the following characteristic condition for the unknown function $\beta(x)$:

$$-x d\beta(x) = \sum_{m=0}^{\infty} d\{\beta^m(x) * \alpha_m(x)\}, \tag{14}$$

where the asterisk denotes the operation of convolution and β^m is an abbreviation for

$$\beta^m = \beta * \cdots * \beta = \beta^{m-1} * \beta. \tag{15}$$

It is understood that, if $m = 0$,

$$\beta^0(x) * \alpha_0(x) = \alpha_0(x) \tag{16}$$

(and that $\beta^1 = \beta$).

It is clear from the definition of the symbols [ ], * that

$$[\lambda * \mu] \leqq [\lambda][\mu]. \tag{17}$$

The integral definition of a convolution also shows that, if either of the by (17). Furthermore, from (20) and (17),

$$[{}^{m+1}\gamma_k] \leqq \sum_{j=1}^{k-1} [\beta_{k-j}][{}^{m}\gamma_j],$$

where [¹ [illegible]], by (19). Since every [ ] is non-negative, these recursive inequalities (in which the members could, for the present, be $+\infty$) imply that the inequalities

$$(23) \qquad [\beta_k] \leqq b_k \qquad (k = 1, 2, \cdots)$$

hold for the sequence $b_1, b_2, \cdots$ defined by the following rule of recursion:

$$(24) \qquad b_{k+1} = \sum_{m=1}^{\infty} [\alpha_m]\, {}^{m}c_k; \qquad b_1 = [\alpha_0],$$

where

$$(25) \qquad {}^{m+1}c_k = \sum_{j=1}^{k-1} b_{k-j}\, {}^{m}c_j; \qquad [illegible]_k = b_k.$$

But the latter system of recursion formulae has a [illegible]ce.

4. In order to see this, put

$$(26) \qquad F(w) = \sum_{m=0}^{\infty} [\alpha_m] w^m$$

and

$$(27) \qquad w = \sum_{k=1}^{\infty} b_k z^k,$$

substitute both (26) and (27) into the equation

$$(28) \qquad w = zF(w)$$

for $w = w(z)$, and compare the coefficients of $z, z^2, z^3, \cdots$ in the power series which thus result on the left and on the right of (28). This supplies for the coefficients, $b_1, b_2, \cdots$, of (27) a recursion formula. But the latter is readily seen to be identical with (24) by virtue of (25).

Hence, the values on the right of the inequalities (23) can be characterized as the coefficients of the solution (27) of (28), provided that (28) has a solution (27). But it does. In fact, since (10), where $[\alpha_m] \geqq 0$, is supposed to hold for some $r > 0$, the power series (26) is the Maclaurin expansion of a function $F(w)$ which is regular at $w = 0$. On the other hand, if $G(z, w) = w - zF(w)$, the partial derivative $G_w(z, w)$ at $(z, w) = (0, 0)$ is $w_w = 1 \neq 0$, and so the equation $G(z, w) = 0$ has a (unique) solution $w = w(z)$ which is regular, and vanishes, at $z = 0$. This means that the equation (28) is satisfied by a (unique) power series (27) which converges functions $\lambda(x), \mu(x)$, where $1 \leqq x < \infty$, vanishes identically for $1 \leqq x < c$, where $c > 1$, then the same is true of the function $\lambda(x) * \mu(x)$. These obvious facts will be essential in what follows.

3. In order to find the unknown function $\beta(x)$, try [illegible]ies,

(18) $$\beta(x) = \sum_{k=1}^{\infty} \beta_k(x).$$

Then (16) and (15) show that what (14) requires is

$$-xd \sum_{k=1}^{\infty} \beta_k(x) = d\alpha_0(x) + d \sum_{m=1}^{\infty} \{ \sum_{k=1}^{\infty} \beta_k(x) * \cdots * \sum_{k=1}^{\infty} \beta_k(x) \} * \alpha_m(x),$$

where $\sum_{k=1}^{\infty} \beta$ [illegible] [illegible]ly m times. Since this means that the m-th { [illegible] where

(19) $$^1\gamma_k(x) = \beta_k(x)$$

and

(20) $$[illegible](x) = \sum_{j=1}^{k-1} {}^m\gamma_j(x) * \beta_{k-j}(x),$$

it follows that what (14) requires of (18) is formally equivalent to the condition

$$\sum_{k=1}^{\infty} -xd\beta_k(x) = d\alpha_0(x) + \sum_{m=1}^{\infty} d \sum_{k=1}^{\infty} {}^m\gamma_k(x) * \alpha_m(x).$$

Clearly, the latter condition is satisfied if (though not only if)

(21) $$-xd\beta_1(x) = d\alpha_0(x)$$

and

(22) $$-xd\beta_{k+1}(x) = d \sum_{m=1}^{\infty} {}^m\gamma_k(x) * \alpha_m(x).$$

This particular disposition is made because (22), (20) and (19) together obviously have the structure of a set of recursion formulae (which commence with (21), the data being the functions α_m). The existence proof can be based on these recursion formulae, as follows:

Since $1 \leqq x < \infty$, it is clear from (3) and (21) that $[\beta_1] \leqq [\alpha_0]$. Similarly, from (22),

$$[\beta_{k+1}] \leqq \sum_{m=1}^{\infty} [\alpha_m][{}^m\gamma_k],$$

near $z = 0$. Incidentally, since the coefficients, (3), of the power series (26) are real and non-negative, it follows, by successive differentiations of (28) at $z = 0$, that $d^k w/dz^k \geqq 0$ holds at $z = 0$ for every k, w[illegible]eans that $b_k \geqq 0$ in (27).

Since (27) has a non-vanishing radius of convergence, [illegible]verges at $z = e^{-l}$ if $l > 0$ is large enough. For such an l,

(29) $$\sum_{k=1}^{\infty} [\beta_k] e^{-kl} < \infty,$$

by (23). This proves, in particular, that

(30) [illegible] $[\beta_k] < \infty$,

i. e., that [illegible] 8) is of bounded variation on the half-line $1 \leqq$ [illegible]

In o[illegible] he proof of the italicized theorem, recourse must be had [illegible] after (17). This remark, when applied to the recursi[illegible] (22) which define $\beta_1, \beta_2, \cdots$, is readily seen to imply t[illegible]

(31) [illegible]$) \equiv 0$ if $1 \leqq x \leqq k$,

(12) being satisfied [illegible]nption) for every $\alpha = \beta_k$.

5. According to (30) and (3), each of the integrals

(32) $${}^k w(s) = \int_1^{\infty} e^{-sx} d\beta_k(x),$$

where $k = 1, 2, \cdots$, is absolutely convergent on the half-plane $\Re s \geqq l$. But (31) shows that (32) can be written in the form

$${}^k w(s) = \int_k^{\infty} e^{-sx} d\beta_k(x).$$

Hence it is clear from (29) and (3) that the series

(33) $$w(s) = \sum_{k=1}^{\infty} {}^k w(s)$$

is absolutely-uniformly convergent on the half-plane $\Re s \geqq l$, and that the same is true if $d\beta_k(x)$ is replaced by $|\, d\beta_k(x) \,|$ in (32), (33).

Accordingly,

$$\sum_{k=1}^{\infty} \int_{1}^{\infty} e^{-lx} \mid d\beta_k(x) \mid < \infty.$$

This implie [illegible] the series (18) is convergent for $1 \leqq x < \infty$, since (12) is satisfie [illegible] β_k. It also follows that, if $\beta(x)$ is defined by (18), and if (32) is substituted into (33), the term-by-term integration leading from (33) to (11) is legitimate on the half-plane $\Re s \geqq l$.

Since the recursion formulae defining the terms of (18) were chosen so as to lead to an integral (11) satisfying (9), the proof of the italicized theorem is now complete, except for the assertion claiming the uniqueness of such a solution. But the uniqueness of [illegible] easy to see. In fact, every function representable, for larg [illegible]s majorized by a constant multiple of exp $(-\Re s)$ [illegible] holds for the derivative of (11). Hence, the uniqueness [illegible]) follows by an obvious adaptation of the standard uniquen [illegible] a (local) Lipschitz constant.

6. Let the derivative of w_i in the differential s [illegible]eplaced by w_i itself. What results, the implicit system [illegible]

(34) $$w_i = f_i(s, w_1, \cdots, w_n), \quad [illegible] \cdots, n),$$

can be treated in the same way as the differential system (6):

The italicized theorem remains true if, in its wording, (34) *is read instead of* (6).

In fact, if (6) is replaced by (34), then what corresponds to (19)-(22) is identical with the recursion formulae which result if the factor $-x$ is suppressed on the left of (21) and of (22). But this factor (which, since $|-x| \geqq 1$, could have been used as divisor promoting convergence) was not used at all. Hence, the above proof contains the proof of the last italicized theorem.

THE JOHNS HOPKINS UNIVERSITY.

ON COMPACT COMPLEX ANALYTIC VARIE[illegible]S.*

BY WEI-LIANG CHOW.

1. Introduction. Let S_n be the projective space of n dimensions over the field k of all complex numbers. For any given choice of the inhomogeneous coordinates $x_1, \cdots, x_n$ in S_n, let $A_n(x)$ be the n-dimensional complex affine space consisting of all the points of S_n which are finite with respect to this coordinate system. [illegible] U in S_n is said to be analytic in the neighborhood of a poi[illegible] inhomogeneous coordinates so that (a) is a point in A[illegible]rhood R of (a) in $A_n(x)$ such that the intersecti[illegible]oints of R which are common solutions of a finite [illegible]

(1) [illegible] $\cdot, x_n) = 0,$ $\qquad (i = 1, \cdots, s),$

where t[illegible]), $i = 1, \cdots, s$, are holomorphic functions in the region R[illegible] S_n is called an analytic variety if it is analytic in the neig[illegible]he of its points. An analytic variety is called compact if it is a comp[illegible]t set. A point set in S_n is called an algebraic variety if it consists of all the points whose homogeneous coordinates $x_0, x_1, \cdots, x_n$ satisfy a finite set of equations

$$\phi_i(x_0, x_1, \cdots, x_n) = 0, \qquad (i = 1, \cdots, t),$$

where the $\phi_i(x_0, x_1, \cdots, x_n)$, $i = 1, \cdots, t$, are homogeneous polynomials in the $x_0, x_1, \cdots, x_n$. It is easily seen that an algebraic variety is a compact analytic variety. There is a classical conjecture that conversely any compact analytic variety in S_n is also an algebraic variety. So far as we are aware, no proof of this conjecture has ever been offered. It is the purpose of this paper to present a proof of this conjecture.

We begin by recalling some well known properties of an analytic variety.[1] Let R be a neighborhood of the origin of $A_n(x)$ defined by the conditions $| x_i | < e$, $i = 1, \cdots, n$. Let $k\{x_1, \cdots, x_r\}$ be the ring of all power series in the variables $x_1, \cdots, x_r$ which are convergent in the region $| x_i | < e$, $i = 1, \cdots, r$. Consider a set of equations of the following type

---

* Received April 9, 1949.

[1] For the results of the theory of functions of several complex variables used in this paper, we refer here once for all to Osgood [7], Bochner and Martin [1].

$$(2)\qquad \begin{aligned} &x_{r+1}^{\mu}+B_1x_{r+1}^{\mu-1}+\cdots+B_{\mu-1}x_{r+1}+B_\mu=0\\ &Hx_j=G_j(x_{r+1}), \qquad (j=r+2,\cdots,n),\end{aligned}$$

where the [illegible] $\cdot, B_\mu$ and H are elements of the ring $k\{x_1,\cdots,x_r\}$ and the G_j [illegible] $r+2,\cdots,n$, are polynomials in x_{r+1} with coefficients in $k\{x_1,\cdots,x_r\}$; and the left hand side of the first equation is an irreducible distinguished polynomial in x_{r+1} over $k\{x_1,\cdots,x_r\}$ and H is its discriminant. Let D be the set of all common solutions of the equations (2) in the region R for which $H\neq 0$, as well as those which are limiting points of such solutions. Such a point set D, or any point set in S_n which can be represented as such a set by a suitable choice of the affine space $A_n(x$ [illegible] coordinates, is called an analytic element in S_n. The number [illegible] ension of the analytic element D, and we shall write [illegible] number μ depends in general upon the choice [illegible] ts coordinates; the smallest possible value of this [illegible] n D_r is called its order. An analytic element is c [illegible] has the order one, otherwise it is called singular. The po [illegible] center of D_r. It is well known from the theory of functi [illegible] complex variables that every point of an analytic variety W [illegible] R such that $R\cap W$ consists of a finite number of anal [illegible] er words, an analytical variety is a topological sub-spac [illegible] which has a system of neighborhoods consisting of analytic elements. An analytic variety is called irreducible if any two analytic elements of it can be obtained from each other by analytic continuations. It is easily seen that all the analytic elements of an irreducible analytic variety W must have the same dimension r; we shall call this number r the dimension of W and we shall write W_r to indicate this. Following a recent practice in algebraic geometry, we shall from now on use the expression "analytic variety" to denote exclusively an irreducible analytic variety; the reducible ones being taken care of later by the more precise concept of an analytic cycle. A point of an analytic variety W_r is called regular if it has a neighborhood consisting of a single regular analytic element; those points of W_r which are not regular are called singular. The set W_r of all singular points of W_r constitutes a finite or enumerably infinite number of analytic varieties of dimensions less than r; we shall call this set $\bar{W}_r$ the singular part of W_r. The set $W_r - W_r$ of all the regular points of W_r is a connected set.

For our present purpose the most important properties of an analytic variety W_r are the following:

(A) *W_r is a topological complex.* This is the triangulation theorem for analytic loci, for the proof of which we shall refer to the literature.[2] In particular, if W_r is compact, it is a finite complex and can be taken as a subcomplex of a simplicial subdivision of the projective sp[illegible]$_n$ and the singular part $\tilde{W}_r$ can in turn be taken as a subcomplex o[illegible]

(B) *W_r can be covered by an enumerable aggregate o*[illegible]*analytic elements of dimensions $\leq r$.* This can be easily proved by induction, starting with the facts that $W_r - \tilde{W}_r$ can be covered by an enumerable aggregate of analytic elements of dimension r and that $\tilde{W}$ consists of an enumerable aggregate of analytic varieties of dimensions $< r$.

In the following sections we shall show that the main theorem can be proved by usin[illegible]rties (A) and (B) alone, without referring directly to [illegible]of W_r has a neighborhood consisting of a finite num[illegible]dimension r. For this reason, we shall give a n[illegible]efinition of an analytic variety in **2**, essentiall[illegible]two properties (A) and (B) only. In **2** and **3** we shall [illegible]l properties of analytic varieties and derive from them th[illegible]hich expresses a special property of a regular analytic element [illegible]a compact analytic variety of r dimensions. In **4** we then pro[illegible] IV [illegible]hat this property is really a characteristic property o[illegible]element D_r which is contained in an algebraic variety of r dimensions[illegible]e main theorem (Theorem V) then follows immediately from these two results. In the last section (**5**) we shall indicate some applications of our main theorem.

Finally, we add a few words about the terminology and notations. A topological complex K is a homeomorphic image of an Euclidean complex, and we shall say $2r$-complex if it consists only of simplexes of topological dimension $2r$ and their sides. The boundary of an unoriented complex K (i. e. considered as a chain modulo 2) will be denoted by $\dot{K}$. If the complex K is oriented to become a chain with integral coefficients, then its chain boundary will also be denoted by $\dot{K}$. There is no danger of confusion in this, as the meaning will be clear from the context in each case. The word "dimension" without any further qualification will mean the complex dimension, which is twice the topological dimension; subscripts will be used to denote the former, superscripts the latter.

2. Analytic varieties. An analytic simplex E_r of r dimensions in S_n is a topological $2r$-simplex in S_n, which is a one-to-one analytic image of a topological $2r$-simplex in a complex affine space $A_r(z)$ of r dimensions. More specifically this means that, for a suitable choice of the inhomogeneous coordinates $x_1, \cdots, x_n$ in S_n, the points of E_r can be represented parametrically by a set of [illegible]tions

---

[2] See Koopman and Brown [4], Lefschetz and Whitehead [6].

$$x_j = f_j(z_1, \cdots, z_r), \qquad j = 1, \cdots, n, \tag{3}$$

where the n functions $f_j(z_1, \cdots, z_r)$, $j = 1, \cdots, n$, are all holomorphic in a region R of the affine space $A_r(z)$ of r complex parameters $z_1, \cdots, z_r$, and the matrix $\| \partial f_j / \partial z_k \|$, $j = 1, \cdots, n$, $k = 1, \cdots, r$, has the maximum rank r at every point of the region R; and that this representation induces a homeomorphism between E_r and a topological $2r$-simplex C^{2r} in the region R. Without loss of generality, we can assume that [illegible] origin $(z) = (0)$ of $A_r(z)$ is contained in the interior of C^{2r} [illegible] $(x) = (a)$ in E_r is then called the center of E_r (wi[illegible] $z_1, \cdots, z_r$). It is obvious that if the analytic s[illegible] complex, then each simplex of this complex is als[illegible] [illegible]ce in all our arguments we can always replace any compl[illegible]n which the simplexes are arbitrarily small; we can there[illegible]ut any essential restriction to the concept of [illegible] analytic sim[illegible] suitable choice of the coordinates $x_1, \cdots, x_n$ [illegible] [illegible]$z)$, the equations (3) have the form

$$\begin{aligned} &x_1 = z_1, \cdots, x_r = z_r \\ &x_{r+1} = f_{r+1}(z_1, \cdots, z_r), \cdots, x_n = f_n(z_1, \cdots, z_r); \end{aligned} \tag{4}$$

and that the region R has been chosen so small that the functions $f_{r+1}, \cdots, f_n$ can be represented as power series which are convergent in R. Since the equations (4) map the region R into a regular analytic element D_r in S_n, it follows that an analytic simplex E_r is simply a topological $2r$-simplex imbedded in a regular analytic element D_r. On the other hand, the set of interior points $E_r - \dot{E}_r$ of an analytic simplex is itself a regular analytic element. Hence it is rather indifferent whether a point set is covered by an aggregate of analytic simplexes or an aggregate of regular analytic elements; the use of either one instead of the other is only a matter of convenience.

The concept of an analytic complex in S_n is defined by induction as follows. Any topological 0-complex in S_n is an analytic complex K_0 of zero dimension. A topological $2r$-complex (finite or infinite) in S_n is called an analytic complex K_r of r dimensions, if it contains a subcomplex $\bar{K}_r$ which consists of an enumerable aggregate of analytic complexes of dimensions less than r, such that any point of $K_r - (\dot{K}_r + \bar{K}_r)$ has a neighborhood which is a regular analytic element D_r. The points of $K_r - (\dot{K}_r + \bar{K}_r)$ are called regular points of K. The topological complex $\dot{K}_r$ is also called the boundary of K_r. The aggregate of analytic complexes $\tilde{K}_r$ is called the singular part of K_r, though not every point (or even any point) of it ne[illegible] a singular point in either the topological or analytic sense. An an[illegible] simplex E_r is of course also an analytic complex of r dimensions, the b[illegible]ing the boundary $\dot{E}_r$ and the singular part being the empty set. An analytic variety W_r of r dimensions in S_n is a connected analytic complex with the properties that it has no boundary and its singular part $\tilde{W}_r$ consists of an enumerable aggregate of analytic varieties of dimensions less than r. This definition by induction can be completed by the stipulation that an analytic variety W_0 of zero dimension consi[illegible]ngle point in S_n. An analytic variety is called compact if it [illegible] that is, if it is a finite topological complex.

It is [illegible] of all the regular points of W_r can be covered b[illegible]f analytic simplexes of r dimensions. From thi[illegible] an enumerable aggregate of analytic varieties [illegible] induction that the analytic variety W_r can b[illegible]te of analytic simplexes of dimensions $\leq$ [illegible] that this covering of W_r by an enumera[illegible]es is not a simplicial subdivision of W_r as a [illegible]ense, nor is it a covering by neighborhoods of W_r as a topo[illegible] space of S_n. It is simply a representation of W_r as the set-theoretic sum of an enumerable aggregate of analytic simplexes of various dimensions, and this aggregate will be in general infinite even in case of a compact analytic variety W_r.

Let E_r and E_s be two analytic simplexes in S_n contained in the regular analytic elements D_r and D_s respectively. The intersection $D_r \cap D_s$ of the two analytic elements D_r and D_s, if it is not empty, consists of an enumerable aggregate of connected components, each of which is an analytic variety of dimension not less than $r + s - n$. In case such a component has exactly the dimension $r + s - n$, it is called a proper component of the intersection. It is easily seen that if the join of the tangent spaces of D_r and D_s at some one regular point of such a component has the dimension n, then the component has the dimension $r + s - n$ and is therefore proper. In such a case we shall call the component regular. We shall say that two analytic simplexes E_r and E_s are regular with respect to each other if either (1) the intersection $E_r \cap E_s$ is empty, or (2) each point of $E_r \cap E_s$ is contained in a regular component of $D_r \cap D_s$.

It is clear that for $r + s < n$ the first possibility alone can occur, so that in this case two analytic simplexes E_r and E_s are regular with respect to each other if and only if they are disjoint. In case $r+s=n$, the intersection of two analytic simplexes E_r and E_{n-r}, regular with respect to each other, is either empt [illegible] consists of a finite number of points at each of which the tangent sp [illegible] D_r and D_{n-r} are transversal to each other. We are mainly interes [illegible] case $r+s=n$, though the following results hold also for the general case.

Let P be the group of all projective transformations in S_n; it is an analytic manifold of $(n+1)^2-1$ dimensions. It is easily seen that if E_r is an analytic simplex in S_n and T is a projective transformation in S_n, then the transformed set TE_r of E_r is also an analytic simplex of r dimensions. It follows then that if K_r is an analytic comp [illegible] by an aggregate of analytic simplexes $\{E\}$, then TK_r is also [illegible] vered by the aggregate of the transformed simplexes [illegible] following theorems, we shall make use of the s [illegible] dense set which is true for every regular sep [illegible] tersection of a finite or enumerable infinite [illegible] n sets is also an everywhere dense set. S [illegible] se if its complement is everywhere dense, it [illegible] enumerably infinite number of nowhere d [illegible] plement of which is everywhere dense; hence [illegible] then it is also nowhere dense.

THEOREM I. *Let E_r and E_s $(r+s\leq n)$ be two analytic simplexes in S_n and let Ω be the set of all elements T of P such that E_r and TE_s are regular with respect to each other, then Ω is an everywhere dense open set in P.*

Proof. We shall choose the inhomogeneous coordinates $x_1,\cdots,x_n$ in S_n in such a way that both E_r and E_s are in the affine space $A_n(x)$, which is always possible if these simplexes are sufficiently small. Let E_r be given by the one-to-one analytic mapping

$$x_i=f_i(z_1,\cdots,z_r), \qquad i=1,\cdots,n, \tag{5}$$

of a region R' in $A_r(z)$ into $A_n(x)$, so that E_r is the image of a topological $2r$-simplex C^{2r} in R'. Similarly, let E_s be given by the one-to-one analytic mapping

$$x_i=g_i(w_1,\cdots,w_s), \qquad i=1,\cdots,n, \tag{6}$$

of a region R'' in $A_s(w)$ into $A_n(x)$, so that E_s is the image of a topological $2s$-simplex C^{2s} in R''.

We first prove that the set Ω is everywhere dense in P; it is evidently enough to show this for the neighborhood of the identity transformation I. Let $T_{(e)}$ be the affine transformation of $A_n(x)$,

$$x'_i = x_i + e_i, \qquad [illegible] \cdots, n;$$

it is of course also a projective transformation in S_n and [illegible]y near to the identity transformation I for points (e) sufficiently near to (0). The transformed analytic simplex $T_{(e)}E_s$ is then given by the one-to-one analytic mapping

$$x_i = g_i(w_1, \cdots, w_s) + e_i, \qquad i = 1, \cdots, n,$$

of the topological simplex C^{2s} in the region R''. Consider the analytic mapping

$$(7) \qquad u_i = [illegible] - g_i(w_1, \cdots, w_s), \qquad i = 1, \cdots, n,$$

of the reg [illegible] space $A_{r+s}(z, w) = A_r(z) \times A_s(w)$ into the [illegible] $(r + s)$-simplex $C^{2r} \times C^{2s}$ in the region R [illegible]olution of the equations (7) for $(u) = ($ [illegible]spond to an intersection point of E_r and [illegible] n, the image of C^{2r+2s} in $A_n(u)$ under the [illegible]s a nowhere dense set; hence there exist points (e) in A_n [illegible] near to the origin such that the equations (7) have no common solution in C^{2r+2s} for $(u) = (e)$. Then the analytic simplexes E_r and $T_{(e)}E_s$ will be disjoint and hence regular with respect to each other. Thus the assertion is proved for this case. For $r + s = n$, we can assume that for all points (e) in a sufficiently small neighborhood of (0) the two analytic simplexes E_r and $T_{(e)}E_s$ have an intersection, for otherwise the assertion is obviously true already. This means that for all points (u) in a sufficiently small neighborhood of the origin in $A_n(u)$ the equations (7) will have a common solution in C^{2r+2s}. It is well known[3] that in this case the Jacobian determinant

$$J(z, w) = | \partial F_i/\partial z_j, \partial F_i/\partial w_k | = | \partial f_i/\partial z_j, -\partial g_i/\partial w_k |,$$

does not vanish identically in the region R. Then the set of all the solutions of the equation $J(z, w) = 0$ in R, if it is not empty, constitutes an enumerable aggregate of analytic varieties of dimension $n - 1$ in R. Let $E^{(i)}$, $i = 1, 2, \cdots$, be a sequence of analytic simplexes covering this set J, and let $E^{(i)*}$, $i = 1, 2, \cdots$, be the corresponding images in $A_n(u)$ under the analytic transformation (7). Then the image J^* of J under the analytic transformation (7) is evidently the sum of all the sets $E^{(i)*}$, $i = 1, 2, \cdots$. Since each $E^{(i)*}$ is a nowhere dense set in $A_n(u)$, it follows that the complement of J^* in $A_n(u)$ i[illegible] everywhere dense set. This means that there exists in every neighborh[illegible] the origin of $A_n(u)$ at least a point (e) which is not in J^*. Then f[illegible]t $(u) = (e)$ the equations (7) will have no common solution in the set J; or, in other words, we have $J(z, w) \neq 0$ at every common solution in R of the equations (7) for $(u) = (e)$. This implies that every intersection point of E_r and $T_{(e)}E_s$ corresponds to a point in C^{2r+2s} at which $J(z, w) \neq 0$; hence E_r and $T_{(e)}E_s$ are regular with respect to each other. Thus our assertion is also proved for the case $r + s = n$.

---

[3] See Knopp and Schmidt [3], p. 379.

It remains to show that the set Ω is open [illegible] is, we have to show that if E_r and E_s are regular with respect [illegible] and TE_s are also regular with respect to each other fo[illegible] l neighborhood of the identity I. For $r + s <$ [illegible] E_r and E[illegible] are disjoint closed subsets in S_n, [illegible] h T. T[illegible] prove the assertion for $r + s = n$ [illegible] be given by the equations

$$x'_i = (a_{i0} + \sum_{j=1}^{n} a_{ij}x_j) / \text{[illegible]} \cdots, n.$$

If the matrix (a_{ij}) is sufficiently near [illegible] the [illegible]ytic simplex TE_s will still lie in the affine space $A_n(x)$ and is given by the one-to-one analytic mapping

$$x_i = g'_i(w_1, \cdots, w_s) = [a_{i0} + \sum_{j=1}^{n} a_{ij}g_j(w_1, \cdots, w_s)]/[a_{00} + \sum_{j=1}^{n} a_{0j}g_j(w_1, \cdots, w_s)], \qquad i = 1, \cdots, n,$$

of the topological simplex C^{2s} in the region R''. Consider the Jacobian determinant

$$J^T(z, w) = | \partial f_i/\partial z_j, \partial g'_i/\partial w_k |,$$

and the n functions

$$F_i^T(z, w) = f_i(z_1, \cdots, z_r) - g'_i(w_1, \cdots, w_s), \qquad i = 1, \cdots, n,$$

which are all defined in the region R. It is clear that these functions are also continuous functions of the transformation T in a neighborhood of the identity I. Now, it is easily seen that the two analytic simplexes E_r and TE_s are regular with respect to each other if and only if the following equations

$$\begin{aligned} &J^T(z, w) = 0, \\ &F_i^T(z, w) = 0, \qquad i = 1, \cdots, n, \end{aligned} \tag{8}$$

have no common solution in the topological simplex C^{2r+2s} in the region R. Let Δ^T be the set of all the common solutions of the equations (8) in R; then Δ^T is closed in R and varies continuously with T. [illegible] E_r and E_s are assumed to be regular with respect to each other [illegible] intersection $C^{2r+2s} \cap \Delta^I$ is empty. Now, if $C^{2r+2s} \cap \Delta^T$ is not empt[illegible] T in a sufficiently small neighborhood of I, then there is a sequence of $T^{(i)}$, $i = 1, 2, \cdots$, converging toward I such that $C^{2r+2s} \cap \Delta^{T^{(i)}}$ is not empty. Since C^{2r+2s} is a compact set, the sequence of sets $C^{2r+2s} \cap \Delta^{T^{(i)}}$ has at least one limit point in C^{2r+2s}; and it is easily seen that this limit point is a point of $C^{2r+2s} \cap \Delta^I$, which is a contradiction. Therefore, the intersection $C^{2r+2s} \cap \Delta^T$ is empt[illegible] T in a sufficiently small neighborhood of the identity I; [illegible] and TE_s are regular with respect to each other for [illegible]s the proof of Theorem I.

Fro[illegible] similar theorem for analytic varieties. We shall [illegible]ersection of two analytic complexes K_r and [illegible]h K_r and K_{n-r} and if the tangent spaces of [illegible]ansversal to each other.

TH[illegible] *two analytic varieties in* S_n, *and let* Ω *be* [illegible]*ns* T *in* P *such that* W_r *and* TW_{n-r} *have only regular* [illegible]*ts. Then the set* Ω *is everywhere dense in* P.

Proof. Let $\{E\}$ and $\{F\}$ be two enumerable sets of analytic simplexes which cover the analytic varieties W_r and W_{n-r} respectively. Then the set $\{G\} = \{(E, F)\}$ of all the pairs of analytic simplexes, one from each of the two sets $\{E\}$ and $\{F\}$, is also enumerable and hence can be arranged in a sequence $G^{(i)} = (E^{(i)}, F^{(i)})$, $i = 1, 2, \cdots$. For each $i = 1, 2, \cdots$, let $\Omega^{(i)}$ be the set of all transformations T in P such that $E^{(i)}$ and $TF^{(i)}$ are regular with respect to each other. It is easily seen that the intersection of all the sets $\Omega^{(i)}$, $i = 1, 2, \cdots$, consists of exactly those transformations T such that W_r and TW_{n-r} have only regular intersection points. Since, according to Theorem I, each set $\Omega^{(i)}$ is an everywhere dense open set in P, it follows that Ω is also an everywhere dense set in P.

3. Intersection of analytic varieties. There is a "natural" orientation for an analytic simplex E_r which can be extended to any analytic complex K_r. Let E_r be an analytic simplex in S_n defined by the one-to-one analytic mapping (3) of the topological $2r$-simplex C^{2r} in $A_r(z)$. Let $z_j = z'_j + iz''_j$, $j = 1, \cdots, r$; then the ordered set of $2r$ real coordinates $z'_1, z''_1, \cdots, z'_r$,

z''_r of the space $A_r(z)$ determines a definite orientation of this space and consequently also an orientation of the simplex C^{2r}. This orientation of C^{2r} determines [illegible]ientation of E_r which we shall call its natural orientation. It can be [illegible]shown that the natural orientation is independent of the choice [illegible]lytic parameters and is thus an intrinsic property of an analytic simplex. The set of all regular points $K_r - (\dot{K}_r + \bar{K}_r)$ of an analytic complex K_r can be covered by a set of analytic simplexes of r dimensions. It can be shown that if each analytic simplex of the set is given its natural orientation, then all these orientations are coherent; and if we orient each topological $2r$-simplex of K_r in concordance with the natural orientation of any one analytic simplex E_r contained in it, we [illegible] topological $2r$-chain on K_r with boundary in the sub-complex [illegible]d $\dot{K}_r$. Thus an analytic complex K_r is an orientable [illegible]s a natural orientation determined by its analytic[illegible] particular also to the projective space S_n its[illegible] complex. Furthermore, if K_r is a finite com[illegible]make it a topological $2r$-chain in S_n with bo[illegible] compact analytic variety W_r, if oriented w[illegible]pological $2r$-cycle in S_n. We shall from no[illegible]omplexes, including the space S_n itself, are or[illegible]tation, so that a finite analytic complex K_r correspo[illegible] a uniquely determined topological $2r$-chain in S_n. The significance of the natural orientation is given by the following well-known result:

LEMMA 1. *The topological intersection multiplicity of a regular intersection point of two finite analytic complexes* K_r *and* K_{n-r} *is* $+1$.

For the proof of this lemma, which is very simple, we refer to the literature.[4]

As to the topological properties of S_n, it is well known that the $2r$-th homology group of S_n is cyclic and the class determined by a linear analytic variety L_r of r dimensions is a generator of this group. From this we can derive at once the following consequences: (1) To each $2r$-cycle in S_n there is associated an integer g called its degree; it is the total topological intersection number of this $2r$-cycle with a linear variety L_{n-r}; (2) two $2r$-cycles are homologous if and only if they have the same degree; (3) the total topological intersection number of a $2r$-cycle of degree g and a $2(n-r)$-cycle of degree h is equal to gh. In view of Lemma 1 we can conclude that the degree of a compact analytic variety W_r is always positive; for it is equal to the number of intersection points of W_r with a linear variety L_{n-r} which has only regular intersections with W_r.

---

[4] For the topological properties of analytic varieties, we refer once for all to Lefschetz [5], Ch. VIII, § 3 and van der Waerden [9].

We shall need the following lemma concerning the [illegible]city of an intersection point. Let B^s and B^{2n-s} be two topological cha[illegible]nd let (p) be an isolated point of $B^s \cap B^{2n-s}$ which does not lie in either $\dot{B}^s$ or $\dot{B}^{2n-s}$. Then this point (p) has a uniquely determined multiplicity as an intersection of the two chains B^s and B^{2n-s}. For the determination of this multiplicity we have the following criterion:

Lemma 2. *Let R be a neighborhood of (p) such that the closure of R does not inters*[illegible] *any other component of the intersection $B^s \cap B^{2n-s}$,* [illegible] *wo chains which are d-homologous to the chains B^s* [illegible] *d is a sufficiently small positive number epending* [illegible] *ach component of the intersection* [illegible]*$^s \cap A^{2n-}$*[illegible] *and the intersection multiplicity of (p) is e*[illegible] *multiplicities of those components of $A^s \cap$* [illegible]

In th[illegible]mology is defined as follows: Let Σ be a d-neig[illegible]od [illegible] closed d-neighborhood of $\dot{B}^s$, then a chain A^s is said to be d-homo[illegible]ous to B^s if it is contained in Σ and is homologous to B^s mod Σ' on Σ. This includes in particular the case when A^s is a d-deformation of B^s. The proof of Lemma 2 follows easily from the well known results about intersection of chains,[5] so that we can omit it here.

Let W_r and W_{n-r} be two compact analytic varieties of degrees g and h respectively, and let (p) be an isolated intersection point of them. Let R be a neighborhood of the point (p) in S_n, and let d be the number such that the Lemma 2 holds for all d-deformations of W_r and W_{n-r}. According to Theorem II, there exists a projective transformation such that TW_{n-r} is a d-deformation of W_{n-r} and the two analytic varieties W_r and TW_{n-r} have only regular intersection points. Since according to Lemma 1 each of these intersection points has exactly the multiplicity $+1$, it follows that the intersection $W_r \cap TW_{n-r}$ consists of exactly gh points. If μ of these gh points lie in the neighborhood R, then μ is the multiplicity of (p) as an intersection of W_r and W_{n-r}. Applying this in particular to a regular point of W_r, we obtain the following result: Let E_r be an analytic simplex in S_n and (p) be an interior point of E_r. If E_r is an element of a compact analytic variety W_r, then there exists a positive number N such that if W_{n-r} is a compact analytic variety of degree h and if (p) is an isolated intersection point of E_r with W_{n-r}, then the multiplicity of this intersection is not greater than hN. ct, we can take N to be any number not less than the degree of W_r. In xt section we shall see that this fact is characteristic of an analyt which is an element of an algebraic variety of r dimensions. However, in order to state this result in the definitive form which we shall use in the next section, we introduce the concept of an analytic cycle. An analytic cycle Z_r of r dimensions in S_n is a topological $2r$-cycle in S_n which can be expressed as a sum of multiples of compact analytic varieties. Thus an analytic cycle Z_r is a finite set of compact analytic varieties $W_r^{(1)}, \cdots, W_r^{(s)}$, to each of which is assigned an integer n_i negative) called its multiplicity. If g_i is the degree of $W_r^{(i)}$ ger $\sum_{i=1}^{s} n_i g_i$ is evidently the degree of the analytic Z_r is calle positive if the multiplicities of its tive; it called algebraic if all of its comp is a simple matter to show that what we hav c varieties can be generalized to positive a following theorem:

---

[5] See Lefschetz [5], Ch. IV, § 3.

THEOREM III. *Let E_r be an an n S_n (p) be an interior point of E_r. If E_r is an element of a compact analytic variety W_r, then there exists a positive number N such that if Z_{n-r} is a positive analytic cycle of degree h and if (p) is an isolated intersection point of E_r with Z_{n-r}, then the multiplicity of this intersection is not greater than hN.*

We have stated the above theorem only for a regular analytic element, as this is the only case we shall need for our purpose. It is however clear that the theorem holds in general for any analytic element, whether regular or not. Incidentally we should like to remark that it is well known[6] that an isolated intersection of two analytic varieties, in the usual sense as defined in section 1, has always a positive multiplicity. This property is fundamental for many applications of topological methods to algebraic geometry, but we have no direct use for it in the present paper. In fact, we do not know whether this property is true at all for the analytic varieties as we have defined here in the previous section; for the proof of it depends on the fact that the neighborhood of each point of an analytic variety consists of a finite number of analytic elements. What is essential for our present purposes is not to rule out the possibility of a zero multiplicity for an isolated intersection point, but to rule out the possibility of a negative multiplicity for any connected component (of any dimension) of the intersection. That the latter is true for intersections of positive analytic cycles in S_n can be easily d[illegible]d from our results, but it is not generally true for intersections o[illegible]posit[illegible]lytic cycles in an arbitrary analytic or algebraic manifold.

---

[6] See van der Waerden [9], or Lefschetz [5], Ch. VIII, § 4.

4. Proof of the main theorem. We begin with two lemmas.

LEMMA 3. *Let $f(x_1, \cdots, x_r)$ be a power series in r variables $x_1, \cdots, x_r$. Given any positive integer N, we can always find an integer M with the following property: For every integer $m > M$, there exists a polynomial $F(x_1, \cdots, x_r, x_{r+1})$ [illegible]iables $x_1, \cdots, x_{r+1}$ of degree m such that the power series [illegible]$, x_r))$ contains no terms of degree $\leq mN$.*

Proof [illegible]he general polynomial of degree m in [illegible]$+ 1$ va[illegible] denotes the binomial coefficient [illegible]$!/j!(h -$[illegible]rms and hence that many indeterminate c[illegible]ies $F(x_1, \cdots, x_r, f(x_1, \cdots, x_r))$ contains a[illegible] mN, the coefficients of which are linear for[illegible]$_r$ linear equations $L_\lambda(c) = 0$ will have a no[illegible]${}^{1+m}\mathscr{C}_{r+1} > {}^{r+mN}\mathscr{C}_r$. This will be the case if we take m gre[illegible]$(N+1)^r(1+r)$; for we have then

$$(m+1) > (1+r)m^r(N+1)^r/m^r$$
$$> (r+1)[(mN+r) \cdots (mN+1)]/[(m+r+1) \cdots (m+2)],$$

hence $[(m+r+1) \cdots (m+1)]/[r+1] > (mN+r) \cdots (mN+1)$, hence ${}^{r+1+m}\mathscr{C}_{r+1} > {}^{r+mN}\mathscr{C}_r$.

Before proceeding to the next lemma we shall recall here some well-known facts about the intersection of two algebraic cycles Z_r and Z_s in S_n. While in general any two arbitrary cycles C' and C'' in S_n have a "topological intersection" consisting of a homologous family of cycles about their geometric intersection $C' \cap C''$, it is possible in case of the algebraic cycles Z_r and Z_s to define an intersection cycle $Z_r \cdot Z_s$ provided all the components of the geometric intersection $Z_r \cap Z_s$ are proper (i. e. have the dimension $r + s - n$). Let $U^{(1)}, \cdots, U^{(\lambda)}$ be the components of $Z_r \cap Z_s$; they are all algebraic varieties of dimension $r + s - n$. Then we can assign in a unique manner to each component $U^{(i)}$ a positive integer n_i such that the algebraic cycle $\sum_{i=1}^{\lambda} n_i U^{(i)}$ is a member of the homologous family of cycles constituting the "topological intersection" of Z_r and Z_s. This cycle $\sum_{i=1}^{\lambda} n_i U^{(i)}$ is called the intersection cycle $Z_r \cdot Z_s$ of Z_r and Z_s, and the numbers $n_1, \cdots, n_\lambda$ are called the intersection multiplicities of $U^{(1)}, \cdots, U^{(\lambda)}$ respectively. The degree of $Z_r \cdot Z_s$ is t[illegible] oduct [illegible] the degrees of Z_r and Z_s. Corresponding to the Lemmas 1 [illegible] o[illegible] preceding section, we have here the following result about [illegible] tion cycle $Z_r \cdot Z_s$: (1) If V_r and V_s are component varieties of multiplicity one of Z_r and Z_s respectively, and if U is a regular component of $V_r \cap V_s$ and is not contained in any other component of Z_r *or* Z_s, then the multiplicity of U in the cycle $Z_r \cdot Z_s$ is equal to $+ 1$. We shall say the U is a regular component of $V_r \cap V_s$ if it contains a point which is regular for both V_r and V_s and if the join of the tangent spaces of V_r and V_s at this point has the dimension n. (2) Give[illegible] itive number d there is a positive number e (depending of cou[illegible] such that if the algebraic cycles Z'_r and Z'_s are e-[illegible] respectively, then the cycle $Z'_r \cdot Z'_s$ is d-homolog[illegible] these ho[illegible] in fact (with slight modificatio[illegible] f analyt cycles and complexes, but we sha[illegible]

LEMMA 4. *Let E_r be an an*[illegible] *t (p) and let L_{n-r+1} be a linear variety con*[illegible] *the intersection of E_r with L_{n-r+1} is an ana*[illegible] *(p). If a positive algebraic cycle Z_{n-1} of degree m*[illegible] *d intersection with E_1 at the center (p) with a multiplicity μ, the*[illegible] *the intersection cycle of Z_{n-1} and L_{n-r+1} is a positive algebraic cycle Z_{n-r} of degree m which has an isolated intersection of multiplicity μ at the point (p) with E_r.*

Proof. The linear variety L_{n-r+1} does not lie in any component hypersurface of Z_{n-1}; for otherwise the entire analytic simplex E_1 would lie in Z_{n-1}, in contradiction to our assumption that Z_{n-1} has an isolated intersection with E_1 at the point (p). Therefore every component of the intersection $L_{n-r+1} \cap Z_{n-1}$ is an algebraic variety of dimension $n - r$ and hence is proper. Therefore the intersection cycle of L_{n-r+1} and Z_{n-1} is a positive algebraic cycle Z_{n-r} of degree m, and it remains only to show that this cycle Z_{n-r} intersects E_r at the point (p) with the multiplicity μ. It is obviously enough to prove the assertion for the case when Z_{n-1} is an irreducible hypersurface, for otherwise we can apply the same argument to each irreducible component of Z_{n-1}. The assumption that Z_{n-1} intersects E_1 at the center (p) with the multiplicity μ implies that there is a projective transformation arbitrarily near to the identity and such that TZ_{n-1} has exactly μ regular intersections with E_1 which lie in any given sufficiently small neighborhood R of (p). For such a T the components of $TZ_{n-1} \cap L_{n-r+1}$ are all proper and hence there is an intersection cycle $Z^T_{n-r} = TZ_{n-1} \cdot L_{n-r+1}$; and since each of the μ regular intersection points of TZ_{n-1} with E_1 is containe [illegible] exactly one regular component of $TZ_{n-1} \cap L_{n-r+1}$, it is contained in e [illegible] o [illegible] mponent of multiplicity one of the cycle Z^T_{n-r}. Let the positi [illegible] be chosen with respect to the chains Z_{n-r} and E_r and the neighborhoo [illegible] point (p) as in Lemma 2, and let the positive number e be chosen with respect to the number d and the intersection cycle $Z_{n-r} = Z_{n-1} \cdot L_{n-r+1}$ as in the remark (2) immediately preceding this Lemma. Since for T sufficiently near to the identity the cycle TZ_{n-1} is an e-deformation of Z_{n-1}, it follows that for such a T the intersection cycle Z^T_{n-r} is d-homologous to Z_{n-r}. It is clear that the part [illegible] ction $Z^T_{n-r} \cap E_r$ which lies in R consists of the same μ i [illegible] Z_{n-1} and E_1. Since each of these points is containe [illegible] t variety of multiplicity one of Z^T_{n-r}, d since [illegible] of this component variety with E_r, erefore [illegible] licity one of Z^T_{n-r} and E_r. It follows th [illegible] plicity of the intersection (p) of Z_{n-r} and E_r [illegible]

With [illegible] can prove the following theorem:

THEO [illegible] *ytic simplex in S_n. Suppose there exists a positive integer N* [illegible] *nat if any positive algebraic cycle Z_{n-r} of degree m in S_n has an isolated intersection with E_r at its center, then the multiplicity of this intersection is at most equal to mN. Then E_r is an element of an algebraic variety V_r in S_n.*

Proof. The theorem is evidently equivalent to the following statement: If E_r is not an element of an algebraic variety V_r in S_n, then given any positive integer N there exists a positive algebraic cycle Z_{n-r} of degree m such that Z_{n-r} has an isolated intersection with E_r of multiplicity greater than mN at the center of E_r.

Let D_r be the analytic element containing E_r. Let the affine space $A_n(x)$ be so chosen that the analytic element D_r is given by the set of equations

$$x_{r+1} = f_{r+1}(x_1, \cdots, x_r), \cdots, x_n = f_n(x_1, \cdots, x_r),$$

where the $f_i(x_1, \cdots, x_r)$, $i = r+1, \cdots, n$, are convergent power series in a region in the affine space $A_r(x)$ of the r complex variables $x_1, \cdots, x_r$. We can assume without any loss of generality that E_r is the image of a simplicial $2r$-simplex C^{2r} in $A_r(x)$ and that the center of E_r is the origin of $A_n(x)$.

Since the analytic simplex E_r, hence also the analytic element D_r is not algebraic, at least one of the $n-r$ power series $x_{r+1}=f_{r+1}(x_1,\cdots,x_r),\cdots,$ $x_n=f_n(x_1,\cdots,x_r)$, [illegible] he power series $x_{r+1}=f_{r+1}(x_1,\cdots,x_r)$, is not an algebraic el[illegible] field $k(x_1,\cdots,x_r)$. According to Lemma 3, there exists to a[illegible] ive integer N a polynomial $F(x)=F(x_1,\cdots,x_{r+1})$ of degr[illegible] in the $r+1$ variables $x_1,\cdots,x_{r+1}$, such that the power series $P(x_1,\cdots,x_r)=F(x_1,\cdots,x_r,f_{r+1}(x_1,\cdots,x_r))$ contains no terms of degree $\leq mN$. Since the element $f_{r+1}(x_1,\cdots,x_r)$ is not algebraic over $k(x_1,\cdots,x_r)$, the power series $P(x_1,\cdots,x_r)$ is not identically zero. Hence, after a suitably chosen linear transformation in $A_r(x)$ if necessary, we can assume that the power series $P(x_1)=P(x_1,0,\cdots,0)$ is not identically zero. It is clear that this power seri[illegible] contains no terms of degree $\leq mN$. Let L_{n-r+1} be the linear [illegible] by the $r-1$ linear equations $x_2=0,\cdots,x_r=0$, th[illegible] ${}_{r+1}$ with D_r is a regular analytic element D_1 give[illegible] $\cdot, x_r=$ [illegible] $x_{r+1}=f_{r+1}(x_1,0,\cdots,0),\cdots,x_n=$ [illegible] tersectio[illegible] of L_{n-r+1} with E_r is an analyt[illegible] which is the image of the topological 2-si[illegible] ntersection with the subspace $x_2=x_3=\cdots=$ [illegible] 2-simplex in the x_1-plane.)

Consider now the homogeneous [illegible]

$$G(x_0,x_1,\cdots,x_{r+1})=x_0{}^mF(x_1/x_0,\cdots,x_{r+1}/x_0)=0;$$

it defines a positive algebraic cycle Z_{n-1} in S_n consisting of as many hypersurfaces as the distinct irreducible factors on the left side of this equation, each taken with its multiplicity. The fact that $P(x_1)=F(x_1,0,\cdots,0,$ $f_{r+1}(x_1,0,\cdots,0))$ vanishes for $x_1=0$ but not identically shows that the point $(x)=(0)$ is an isolated intersection of Z_{n-1} and E_1. It only remains to prove that the multiplicity of this intersection is greater than mN, for then our theorem will follow from Lemma 4. Here again we can assume that the algebraic cycle Z_{n-1} is an irreducible hypersurface, which means that the form $G(x_0,x_1,\cdots,x_{r+1})$ and hence also the polynomial $F(x_1,\cdots,x_{r+1})$ is irreducible; for otherwise we can apply the same argument to each component hypersurface of Z_{n-1}. According to Lemma 2, it is sufficient to show that given any neighborhood R of the point $(x)=(0)$ in $A_n(x)$ there exists a linear transformation $T_{(e)}$,

$$x'_i=x_i+e_i, \qquad i=1,\cdots,n$$

arbitrarily near to the identity, such that $T_{(e)}Z_{n-1}$ has more than mN regular intersections with E_1 in R.

Consider the power series in x_1 with coefficients in $k(u) = k(u_1, \cdots, u_{r+1})$

$$\Phi(x_1, (u)) = F(x_1 + u_1, u_2, \cdots, u_r, u_{r+1} + \text{[illegible]}, 0, \cdots, 0)),$$

which is convergent for all sufficiently small x_1 [illegible] (u). The intersection of $T_{(e)}Z_{n-1}$ and E_1 is then given by the zer[illegible] [illegible]er series $\Phi(x_1, (-e))$ in C^2, and it is well known that a simple zero $x_1 = b_1$ corresponds to a point $(b) = (b_1, 0, \cdots, 0, f_{r+1}(b_1, 0, \cdots, 0), \cdots, f_n(b_1, 0, \cdots, 0))$ on E_1 which is a regular intersection of $T_{(e)}Z_{n-1}$ and E_1. Now we have the fact that $\Phi(x_1, (0)) = P(x_1)$ has a zero of order $\mu > mN$ at the point $x_1 = 0$. Hence, by the Weierstrass Preparation Theorem, there is a neighborhood R' of $x_1 = 0$ and a neighborhood R'' of $(u) = (0)$ such that

$$\text{[illegible]}q(x_1, (u))Q(x_1, (u)),$$

where $Q($[illegible] [illegible]olynomial of degree μ in x_1 with center at $x_1 = 0$ [illegible] a power series which does not vanish for x_1 in [illegible] [illegible]shed polynomial $Q(x_1, (u))$ might be reducib[illegible] [illegible]e field of all quotients of power series in [illegible] [illegible]ctors. In fact, if any power series $q(x_1, (u)$[illegible] is a multiple factor of $Q(x_1, (u))$, then the [illegible] [illegible]s a multiple factor of $Q(0, (u))$ and hence also a multiple fa[illegible] the polynomial $\Phi(0, (u)) = F(u)$. One observes that the power series $q(u)$ is not a constant, for $q(u)$ is not identically zero and we have $q(0) = 0$. Hence the equation $q(u) = 0$ defines a finite number of analytic elements of r dimensions in $A_{r+1}(u)$ with center at $(u) = (0)$. If one of the variables (u), say u_1, is actually involved in the polynomial $F(u)$, then $\partial F(u)/\partial u_1$ is also a polynomial in (u). It follows then that both polynomials $F(u)$ and $\partial F(u)/\partial u_1$ are divisible by the power series $q(u)$. This means that the two positive algebraic cycles of r dimensions in $A_{r+1}(u)$ defined by the two equations $F(u) = 0$ and $\partial F(u)/\partial u_1 = 0$ have at least an analytic element of r dimensions in common and consequently also at least one component variety in common. Since $F(u)$ is an irreducible polynomial, the algebraic cycle $F(u) = 0$ is an irreducible algebraic variety and hence must be a component variety of the algebraic cycle $\partial F(u)/\partial u_1 = 0$. This means that the polynomial $F(u)$ is a factor of the polynomial $\partial F(u)/\partial u_1$, both being now considered as elements of the polynomial ring $k[u]$. But this is impossible, for the polynomial $\partial F(u)/\partial u_1$ has a lower degree than $F(u)$. Thus we have shown that the distinguished polynomial $Q(x_1, (u))$ has no multiple factors.[7]

---

[7] The last part of the argument consists essentially in showing that an irreducible

Since $Q(x_1, (u))$ has no multiple factors, its discriminant $H(u)$ is not identically zero in R''. nce for any point (e) in R'' such that $H(e) \neq 0$, the polynom Q(has exactly μ simple zeros and consequently also the power s) has exactly μ simple zeros in R'. For all points (e) suff to the point $(u) = (0)$ with $H(e) \neq 0$, the μ simple zeros will lie in any given neighborhood of the point $x_1 = 0$ and hence the corresponding μ intersection points on E_1 will lie in any given neighborhood of the point $(x) = (0)$. This completes the proof of the theorem.

Combining Theorem III and Theorem IV, we obtain immediately the main theorem:

THEOREM V. *A compact analytic variet s algebraic variety.*

Proof. Let E_r be an analytic simple ct algebraic variety W_r. According to Theorem III, algebraic variety V_r which also contains the inciple o, analytic continuation, any two (ir eties of dimensions which have an analy coincide. Hence we have $W_r = V_r$, and the

It is perhaps not without inte between Theorem IV and the well known cri braic by means of diophantine approximation: A number ebraic if and only if there exists a positive integer N such that no form $\sum_{i=0}^{n} z_i \xi^i$ has a proper approximation t^{-N} with respect to the variables $z_0, z_1, \cdots, z_n$. In other words, if $Z = \max(|z_0|, |z_1|, \cdots, |z_n|)$, then the number ξ is algebraic if and only if for each n the diophantine inequality

$$\left| \sum_{i=0}^{n} z_i \xi^i \right| < Z^{-N}$$

has only a finite number of solutions for which the left hand side is not zero. Hence, for each n there exists a positive number Γ_n such that the diophantine inequality

$$\left| \sum_{i=0}^{n} z_i \xi^i \right| < (\Gamma_n Z)^{-N}$$

has no solution for which the left hand side is not zero. Taking the negative of the logarithm of both sides, we can write the inequality as follows:

---

polynomial $F(u_1, \cdots, u_{r+1})$ cannot have multiple factors in the ring of power series $k\{u_1, \cdots, u_{r+1}\}$. This is a special case of a theorem proved by Chevalley, see [2], p. 11, Theorem 1, which asserts that any prime ideal in $k(u_1, \cdots, u_{r+1})$ is the intersection of prime ideals in $k\{u_1, \cdots, u_{r+1}\}$.

$$-\log \left| \sum_{i=0}^{n} z_i \xi^i \right| > N(\log Z + 1\text{[illegible]}).$$

In the analogy between algebraic number field an [illegible] nction field of one variable we can regard the number ξ as an "a [illegible]" and the equation $\sum_{i=0}^{n} z_i y^i = 0$, for a given (z), as an "algebraic cur [illegible] The number $\log Z + \log \Gamma_n$ can then be regarded as representing the "degree" of the "algebraic curve" and the left hand side of the above inequality as the "intersection multiplicity" of the "analytic branch" with the "algebraic curve." Carrying this over to the algebraic function field of one variable, we obtain the statement [illegible] analytic branch E_1 in S_2 is algebraic if and only if there [illegible] that no algebraic curve of degree m can have an iso [illegible] with a multiplicity greater than mN. [illegible] hus our [illegible] ed as the extension of this criterion [illegible] analytic [illegible] nensions.

Finall [illegible] t though Theorem IV is stated and proved [illegible] ent, it is not difficult to see that it holds [illegible] hether regular or singular. Thus this theore [illegible] s both necessary and sufficient for any analyti [illegible] ent to b [illegible] Though our proof is partly algebraic and partly topological, as i [illegible] cessary for our present purpose, the theorem itself can be expressed in purely algebraic terms, using only formal power series. In fact, one could probably obtain a purely algebraic proof of this result by means of the intersection theory for algebroid varieties as developed by Chevalley.[8]

5. Meromorphic transformations. We begin with an almost obvious generalization of Theorem V to a multiply projective space. For the sake of convenience we shall restrict ourselves to the case of a doubly projective space $S_m \times S_n$, though the results hold obviously for the general case. We can define an analytic variety W_r in $S_m \times S_n$ in exactly the same way as we have done in **2** for S_n. Now the space $S_m \times S_n$ can be mapped by a bi-regular birational transformation onto a non-singular algebraic variety V_{m+n} in a suitably chosen projective space S_T. Under this transformation any analytic or algebraic variety in $S_m \times S_n$ is carried into an analytic or algebraic variety in V_{m+n} respectively, and conversely. Therefore, if W_r is a compact analytic variety in $S_m \times S_n$, then its image W_r^* is also a compact analytic variety in

---

[8] Chevalley [2].

V_{m+n} and hence also in S[illegible] By Theorem V, $W_r{}^*$ is an algebraic variety, hence W_r is also an algebra[illegible]iety.

THEOR[illegible]*mpact analytic variety in a multiply projective space is an alge*[illegible].

In fact, we can extend this theorem to all the so-called "extended spaces" of Osgood,[9] but we shall not go into details here. Instead we shall derive from Theorem VI an important result concerning meromorphic transformation of an analytic variety. A mapping of an analytic variety W_r in S_n into the space S_m is called a meromorphic transformation if in a sufficiently small neighborhood R of any point (a) of W_r and fo[illegible]uitable choice of the affine spaces $A_n(x)$ and $A_m(y)$ in S_n and S_m re[illegible]ing is defined by a set of equations

$$(9) \qquad g(x)y_i = f_i(x \quad [illegible] \quad \cdots, m,$$

where the functions $g(x), f_1(x), \cdot$ [illegible] and $g(x)$ does not vanish for all points of [illegible] to be understood with the following stipulat[illegible] $W_r \cap R$ such that $g(p) \neq 0$, the image unde[illegible] $f_i(p)/g(p)$, $i = 1, \cdots, m$, in S_m. We shall call [illegible] (p) a reg[illegible] point of the mapping. For a point $(x) = (p)$ such that [illegible] $= 0$, the image under the mapping is to be the set of all points in S_n which are limits of the images of a sequence of regular points approaching (p). Thus a meromorphic transformation, just like the rational transformation in case of an algebraic variety, is not a mapping in the strict sense of the word; it assigns only to a sufficiently "general" point of W_r a unique image, while to the others (which correspond to the fundamental points of a rational transformation) it assigns certain subsets as images.

THEOREM VII. *A meromorphic transformation of a compact analytic (hence algebraic) variety W_r in S_n into a projective space S_m is a rational transformation of W_r, and the image is an algebraic variety in S_m.*

Proof. The graph of the meromorphic transformation in the product space $S_n \times S_m$ is a point set G, the projection of which in S_n is the variety W_r. Let $((a), (b))$ be any point of G, and let R be a sufficiently small neighborhood of $((a), (b))$ with R' and R'' as its projections in S_n and S_m respectively. The variety W_r being algebraic, it is defined in the neighborhood

---

[9] See Osgood [7], Ch. III, § 32.

R' of the point (a) by a set of algebraic equations [illegible] $(x) = 0$, $i = 1, \cdots, s$. Then the set $G \cap R$ will be defined by the set of [illegible] ons

$$
(10) \qquad \begin{aligned} g(x) y_i &= f_i(x), \qquad [illegible] \cdots, m, \\ \phi_j(x) &= 0, \qquad [illegible] \cdots, s, \end{aligned}
$$

with the stipulation that any component variety lying entirely in the hypersurface $g(x) = 0$, if it is an isolated one, should be deleted from the set. It is well known[10] from the theory of power series ideals that this can be achieved by the adjunction of a finite number of suitably chosen analytic equations to the equations (10), if the neighborhood R is chosen sufficiently small. Thus the set $G \cap$ [illegible] consists of all the common solutions of a set of a [illegible] Since $((a), (b))$ is any point of G, this means that [illegible] in $S_n \times S_m$.

It re [illegible] pact set. Let $((a^{(i)}), (b^{(i)}))$, $i = 1$, $\cdots$, be [illegible] verging to a point $((a), (b))$ in $S_n \times S_m$; w [illegible]) is a point of G. Since W_r is compact and [illegible] int of the sequence $(a^{(i)})$, $i = 1, 2, \cdots$, [illegible] In a sufficiently small neighborhood of [illegible] sformation will be represented by a set of equ [illegible] ns (9), w [illegible] sponding stipulations. We can assume without loss of generality [illegible] for all large i the points $(a^{(i)})$ are regular, i. e. $g(a^{(i)}) \neq 0$. For, since each point $((a^{(i)}), (b^{(i)}))$, with $g(a^{(i)}) = 0$ is by our stipulation the limit of a sequence of regular points, we can always replace each point $((a^{(i)}), (b^{(i)})$, for sufficiently large i, by a regular point sufficiently near to it so that the resulting sequence will have the same limit point $((a), (b))$. Now, if $g(a) \neq 0$, then we have

$$
b_j = \lim_{i \to \infty} b_j^{(i)} = \lim_{i \to \infty} f_j(a^{(i)})/g(a^{(i)}) = f_j(a)/g(a), \qquad j = 1, \cdots, m;
$$

and hence $((a), (b))$ is a point of G. On the other hand, if $g(a) = 0$, then the point (b) is by our stipulation a point of G. Therefore G is a compact point set.

Thus we have shown that G is a compact analytic variety in $S_n \times S_m$; hence, by Theorem VI, G is an algebraic variety. It follows then that the projection of G in S_m is also an algebraic variety W^* and the given meromorphic transformation is really an (irreducible) algebraic correspondence between W_r and W^*. Since this algebraic correspondence assigns to a generic point of W_r a unique image point in W^* (and since our ground field has the characteristic 0), [illegible]fore it is a rational transformation of W_r onto W^*. Thus the proof is co[illegible]d.

---

[10] See Rückert [8], or Bochner and Martin [1], Ch. X.

The co[illegible]romorphic transformation evidently includes that of a merom[illegible]on as the special case $m = 1$. Hence we have the corollar[illegible]

Corollary. *An everywhere meromorphic function on a compact analytic variety is a rational function.*

So far as we know, even this corollary has been proved up to now only for a few special cases such as the projective space and the space of analysis (the product of projective lines).

The Johns Hopkins University

RE[illegible]

1. S. Bochner and W. T. Martin, *Sever*[illegible]o.
2. C. Chevalley, "Intersections of algebr[illegible] variet[illegible]*nsactions of the American Mathematical Society*, [illegible]5), pp. 1-85.
3. K. Knopp and R. Schmidt, "Funktionaldetermina[illegible] und Abhaengigkeit von Funktionen," *Mathematische Zeitschrift*, vol. 25 (1926), pp. 373-381.
4. B. O. Koopman and A. B. Brown, "On the covering of analytic loci by complexes," *Transactions of the American Mathematical Society*, vol. 34 (1932), pp. 231-251.
5. S. Lefschetz, *Topology*, New York, 1930.
6. ——— and J. H. C. Whitehead, "On analytical complexes," *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 510-517.
7. W. F. Osgood, *Lehrbuch der Funktionentheorie*, Bd. II, 1. Leipzig, 1929.
8. W. Rückert, "Zum Eliminationsproblem der Potenzreihenideale," *Mathematische Annalen*, vol. 107 (1932), pp. 259-281.
9. B. L. van der Waerden, "Topologische Begründung des Kalküls der abzählenden Geometrie," *Mathematische Annalen*, vol. 102 (1929), pp. 337-362.

A CHARACTERIZATION OF THE SP A OF ONE-DIMENSIONAL WAVE EQUA

By PHILIP HARTMAN.

---

1. Let a positive $p(t)$ and a real-valued $q(t)$, where $0 \leqq t < \infty$, be continuous functions with the property that the differential equation

$$(1) \qquad (px')' + (q + \lambda)x = 0$$

is in the Gre eans ([4], p. 238) that, for some (and then for every) which is not of class $L^2(0, \infty)$. Thus 1) and a

$$(2) \qquad = 0 \qquad (0 \leqq \alpha < \pi)$$

determin Hilbert space $L^2(0, \infty)$. Let $S(\alpha)$ denote lem (1), (2).

Let $x($ be a lution of (1), (2) and let $N(T, \lambda)$ denote the number of it $\leqq t < T$. Finally, for $\lambda' < \lambda''$, let $n = n(\lambda', \lambda'') \leqq \infty$ denote limit inferior, as $T \rightarrow \infty$, of the difference

$$(3) \qquad N(T, \lambda'') - N(T, \lambda').$$

The following theorem will be proved:

(I) *There are exactly n points* λ *of* $S(\alpha)$ *satisfying* $\lambda' < \lambda < \lambda''$ *or* $\lambda' \leqq \lambda < \lambda''$ *according as* (1) *is oscillatory or non-oscillatory for* $\lambda = \lambda''$; *in particular,* $n = \infty$ *if and only if an infinite set of points* λ *of* $S(\alpha)$ *satisfies* $\lambda' < \lambda < \lambda''$.

What is essential and new in (I) is the fact that no restriction is placed on (1), except that it be in the *Grenzpunktfall.* In particular, it is not assumed that (1) is non-oscillatory for λ on $\lambda' \leqq \lambda \leqq \lambda''$ as in earlier theorems, or even that (1) is non-oscillatory for some λ. Actually, the oscillatory case of (1) has always been the more complicated in view of the comparatively simple structure of the solutions of (1) in the non-oscillatory case; cf., e. g., [1], pp. 701-703. For this reason, the proofs for the non-oscillatory cases cannot be adapted to (I). Oscillation theorems of the type (I) go back to

---

* Received May 5, 1949.

Weyl [4], pp. 251-2[illegible] here it is assumed that (1) is non-oscillatory for $\lambda' \leqq \lambda \leqq \lambda''$ and th[illegible] $\geqq$ const. The author has shown [1] that Weyl's method can [illegible] as to apply to the general non-oscillatory case without any as[illegible](t), except that (1) be in the *Grenzpunktfall.*

T[illegible]) is similar to those recently used in [3] in connection with separatio[illegible]eorems for the spectra of Hermitian matrices and their sections; in particular, see the last part of (iii) in the Appendix and the proof of theorem (*) in [3]. The assertion (I) and its proof will show and will depend on the fact that the spectrum $S(\alpha)$ can be obtained as the "limit" of the spectra $S_T(\alpha)$ of Sturm-Liouville boundary value problems on $0 \leqq t \leqq T$, as $T \to \infty$. This answers a questio[illegible]aised by Wintner regarding the definition and validity of the limit process $S_T(\alpha)$ [illegible]

Let $\mu_0(T) < \mu_1(T) < \cdots$ denote [illegible]alues of the Sturm-Liouville boundary value prob[illegible]ned by (1[illegible] the boundary conditions (2) at t[illegible]

(4) [illegible]

at $t = T$.

(II) *A number* λ *belongs to* S[illegible] *if* $d($[illegible]*, as* $T \to \infty$, *where* $d(T) = \min | \lambda - \mu_j(T) |$ *for* $j = 0, 1, \cdots$.

The point in (II) is that in order that λ be in $S(\alpha)$, it is not sufficient that $d(T) \to 0$ be true merely on a sequence of T-values, which tend to ∞. The statement of (II) gives an "elementary" characterization of points of the spectrum $S(\alpha)$ of (1), (2).

Let $\epsilon > 0$ and $T = T_\epsilon$ be so large that, for some value of λ, $d(T) = d(T, \lambda)$ satisfies $| d(T) | \leqq \epsilon$ whenever $T > T_\epsilon$. Then there is at least one eigenvalue of $S_T(\alpha)$, say $\mu_h(T)$, in $[\lambda - \epsilon, \lambda + \epsilon]$ for $T > T_\epsilon$. Actually, for some arbitrarily large values of T, there are at least two points $\mu_h(T)$, $\mu_{h+1}(t)$ in $[\lambda - \epsilon, \lambda + \epsilon]$ in the oscillatory case. For consider a T-value $T = T' > T_\epsilon$ for which $x(T', \lambda - \epsilon) = 0$, so that $\lambda - \epsilon = \mu_h(T')$ for some h. Then $\mu_{h+1}(T') \leqq \lambda + \epsilon$, for otherwise for some values of T near but greater than T, it follows that $d(T) > \epsilon$, since $\mu_h(T) < \lambda - \epsilon$ and $\mu_{h+1}(T) > \lambda + \epsilon$. This fact (involving μ_h, μ_{h+1}) shows that assertion (II), above, is an analogue of the theorem (iii) in the Appendix of [3].

The boundary condition (4) in (II) can be replaced by any fixed homogeneous boundary condition

$$x(T) \cos \beta + x'(T) \sin \beta = 0, \qquad (0 \leqq \beta < \pi), \tag{4 bis}$$

at $t = T$. In fact, (1) can be modified as follows: $' < \lambda''$, let m denote the limit inferior, as $T \rightarrow \infty$, of the number of eig $\mu_0(T), \mu_1(T), \cdots$ belonging to the boundary value problem (1), (2 the interval $\lambda' \leqq \lambda \leqq \lambda''$, then (I) remains true if "m" is read ."

It may be remarked that the difference between the nd non-oscillatory cases of (I) is partially explained by the fa t in the latter case $n = n(\lambda', \lambda'')$ is not only a "limit inferior," but act ly a "limit." Thus if (1) is non-oscillatory for $\lambda = \lambda''$, then $n(\lambda', \lambda) + n(\lambda, \lambda'') = n(\lambda', \lambda'')$ holds for $\lambda' < \lambda < \lambda''$, as a consequence of $\{N(T, \lambda'') - N(T, \lambda)\} + \{N(T, \lambda) - N(T, \lambda')\} = N(T, \lambda'') - N(T, \lambda')$; in particular, if $\lambda = \lambda''$ is an isolated point of $S(\alpha)$, then $n(\lambda'' - \epsilon, \lambda'')$ 0 and $n(\lambda'', \lambda'' + \epsilon) = 1$, while $n(\lambda'' - \epsilon, \lambda'' +$ 1 ll $\epsilon > 0$. On the other hand, if (1) is oscillatory for $\lambda = \lambda''$ isolated point of $S(\alpha)$, then (I) implies $(\lambda'' - \epsilon, \lambda''$ $= 0$, but $n(\lambda'' - \epsilon, \lambda'' + \epsilon) = 1$ for mall $\epsilon > 0$, $n = n(\lambda', \lambda'')$ is not additive.

imp theorem, known for the case when (1) is p

for $\lambda' < \lambda < \lambda''$. *Then* (i) $N(\lambda)$ *is a non-dec (int ep function*; (ii) $N(\lambda)$ *is continuous from the left*; (iii) *eve (if any) of* $N(\lambda)$ *has the value* 1; *finally,* (iv) *the discontinuity po f* $N(\lambda)$ *on* $\lambda' < \lambda < \lambda''$ *are in the point spectrum and no other* λ*-values satisfying* $\lambda' < \lambda < \lambda''$ *are in the spectrum* $S(\alpha)$ *of* (1), (2).

2. *Proof of* (I). Associated with (1), (2) is a self-adjoint operator $L(z) = (pz')' + qz$ defined on the set of functions $z(t)$ for which z, pz' are of class C' on $0 \leqq t < \infty$; and z, $L(z)$ are of class $L^2(0, \infty)$; finally, $x = z$ satisfies (2) at $t = 0$. The set $S(\alpha)$ of λ-values is the spectrum of this self-adjoint operator.

In order to avoid the consideration of different cases and duplication of known theorems, the proof of (I) will be given in the case that (1) is oscillatory for $\lambda = \lambda'$. For the case that (1) is non-oscillatory for $\lambda = \lambda''$, see [1]; for the case that (1) is non-oscillatory for $\lambda = \lambda'$ but oscillatory for $\lambda = \lambda''$, see [2]. (Actually, the arguments below with slight modifications give simpler proofs for these cases also.)

It will first be shown that $S(\alpha)$ contains at least n points of $\lambda' < \lambda < \lambda''$. If $n < \infty$, it is possible to choose T so large that (3) has the value n and that $x = x(t, \lambda')$ satisfies (4), since (3) is not increasing at T when (4) holds for $x = x(t, \lambda')$. If $n = \infty$, a number T can be chosen so large that the value of (3), which will b[illegible] n for the moment, exceeds any given number and that $x=x(t,\lambda')$ s[illegible] (4). It may also be supposed that $x=x(t,\lambda'')$ satisfies (4) [illegible] can be replaced by a smaller value of λ, say $\lambda=\lambda''$, for which [illegible] $=x(t,\lambda'')$ and (3) has the value n.

Le[illegible] $T) < \cdots$ denote the eigenvalues of the Sturm-Liouville boundary valu[illegible]lem on $0 \leqq t \leqq T$ determined by (1), (2), (4), and let $x=x_j(t)$, w[illegible] $j=0,1,\cdots$, denote a corresponding set of eigenfunctions on $0 \leqq t \leqq T$. Since $x=x(t,\lambda')$ and $x=x(t,\lambda'')$ satisfy (2), (4), and since (3) has the value n, it follows that $\lambda'=\mu_h$ and $\lambda''=\mu_{h+n}$ for some h. Put

$$z(t)=\sum_{j=0}^{n} c_j x_{h}\,[illegible]\,(t) \quad \text{or} \quad z(t)=0 \tag{5}$$

according as $0 \leqq t \leqq T$ or $T < t < \infty$, w[illegible]re arbitrary constants. An appropriate choice of these [illegible]below.

Suppose, if possible, that only [illegible]e contained i[illegible] $S(\alpha)$, and that $k < n$. Let λ_1 [illegible] k v[illegible] (if $k > 0$). These λ-values are in t[illegible] equ[illegible]ons

$$\int_0^{\infty} x(t,\lambda_j)z(t)\,dt \; [illegible] \int_0 [illegible]\,dt \tag{6}$$

for $j=1,2,\cdots,k$, and the equation

$$z'(T-0)=0 \tag{7}$$

represent at most $k+1$ homogeneous linear equations for the $n+1$ constants $c_0, c_1,\cdots,c_n$. Since $k+1 < n+1$, these equations possess a non-trivial solution. In (5), let $c_0, c_1,\cdots,c_n$ denote such a solution of (6), (7), so that $z(t) \not\equiv 0$.

Since $x=x_j(t)$ satisfies (4), it follows from (1), for $\lambda=\mu_j$, that $(px_j')'=0$ at $t=T$. Hence (5) and (7) show that $z=z'=(pz')'=0$ at $t=T$ and that z and pz' are of class C' for $0 \leqq t < \infty$. Also, the second of the conditions (5) shows that z, $L(z)$ are of class $L^2(0,\infty)$. Finally, $x=z$ satisfies (2) at $t=0$, since $x=x_j$ does. Hence, $z=z(t)$ is in the domain of definition of the self-adjoint operator $L(z)$. From (1) and (5), it is seen that

$$L(z)+\tfrac{1}{2}(\lambda'+\lambda'')z=\sum_{j=0}^{n} c_j\{\tfrac{1}{2}(\lambda'+\lambda'')-\mu_{h+j}\}x_{h+j}$$

for $0 \leqq t \leqq T$ and that $L(z)+\tfrac{1}{2}(\lambda'+\lambda'')z=0$ for $T < t < \infty$. Since $|\tfrac{1}{2}(\lambda'+\lambda'')-\mu_{h+j}| \leqq \tfrac{1}{2}(\lambda''-\lambda')$ for $j=0,1,\cdots,n$, it follows from (5) that

$$\int_0^\infty \{L(z)+\tfrac{1}{2}(\lambda'+\lambda'')z\}^2 dt \leqq [\tfrac{1}{2}(\lambda''- \qquad {}^{\infty} z^2(t)\,dt, \tag{8}$$

in view of the orthogonality of the eigenfunctions x_{h+n}
$0 \leqq t \leqq T$.

However, since (6) means that $z(t)$ is orthogonal to finite set of eigenfunctions belonging to eigenvalues of $L(z)$ within a distance $\tfrac{1}{2}(\lambda''-\lambda')$ of the point $\tfrac{1}{2}(\lambda'+\lambda'')$, that is, from within $\lambda' < \lambda < \lambda''$, it follows from the Parseval identity belonging to the spectral resolution of $L(z)$ that the sign of inequality cannot hold in (8). For the same reason, the sign of equality can hold if and only if at least one of t numbers λ', λ'' is an eigenvalue and
$z(t)$ is a linear combina cor ponding functions $x(t,\lambda')$, $x(t,\lambda'')$.
But this is impo a t of (5). Thus the assumption that
$< \lambda < \lambda''$ co f $S(\alpha)$ leads to a contradiction. This
oves that at $\lambda' < \lambda < \lambda''$.
$n <$ t most n points of $S(\alpha)$ are in
$\lambda' <$, (3) has the value n when (4)
holds e exists an $h = h(T)$ such that
$\lambda' = \mu_h(T) <$ $\lambda'' < \mu_{h+n+1}(T)$. Hence, if (4)
holds for $x = x(t,\lambda')$, the $\lambda \leqq \lambda''$ contains exactly n eigenvalues
$\mu_{h+1}, \cdots, \mu_{h+n}$ of the Stu aville boundary value problem (1), (2), (4).

Choose a sequence of lues, $T_1, T_2, \cdots$, which tend to ∞ in such a way that, as $i \to \infty$, $\lim \mu_{h+j}(T_i) = \mu^j$ exist for $j = 1, 2, \cdots, n$. It will be shown that the points of $S(\alpha)$ in $\lambda' < \lambda < \lambda''$ are contained in the set of (at most n) points $\mu^1, \cdots, \mu^n$. Let λ be a point in $\lambda' < \lambda < \lambda''$ distinct from $\mu^1, \cdots, \mu^n$. Let $d > 0$ be so small that

$$\min\,(\lambda''-\lambda,\ \ \lambda-\lambda',\ |\lambda-\mu_{h+j}(T)|) > d > 0$$

holds for all large $T = T_i$ and for $j = 1, \cdots, n$. Finally, let $g = g(t)$, where $0 \leqq t < \infty$, be an arbitrary continuous function of class $L^2(0,\infty)$.

The last set of inequalities implies that if $T = T_i$ is sufficiently large, then

$$(px')' + (q+\lambda)x = g \tag{9}$$

has a solution $x = x_T(t)$ satisfying (2), (4); furthermore,

$$\int_0^T x_T{}^2 dt \leqq d^{-2}\int_0^T g^2 dt \leqq d^{-2}\int_0^\infty g^2 dt,$$

since the distance from λ to the nearest of the eigenvalues $\mu_0(T), \mu_1(T), \cdots$

exceeds d. Clearl uence $T_1, T_2, \cdots$ contains a subsequence $R_1, R_2, \cdots$ such that, as $x_R(t) = x(t)$, where $R = R_i$, exists (uniformly on every finit $t \leq T$). Hence $x(t)$ is a solution of (9) satisfying) and

$$\int_0^\infty x^2 dt \leqq d^{-2} \int_0^\infty g^2 dt < \infty.$$

Since g is an arbitrary continuous function of class $L^2(0, \infty)$, it follows that λ does not belong to $S(\alpha)$; [4], p. 251. This completes the proof of (I).

THE JOHNS HOPKINS UNIVERSITY.

R

[1] P. Hartman, "Differential equa Duke *Mathematical Journal*, vol.

[2] ——— and C. R. Putnam, "The lea rum of boundary value problems," *American Journ matics*, vol. 70 (1948), pp. 849-855.

[3] ——— and A. Wintner, "Separation theore bounded Hermitian forms," *ib l.*, vol. 71 (1949), pp. 865-878.

[4] H. We "Ueber gewöhnliches Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlichen Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 220-269.

THE JOHNS HOPKINS UNIVERSITY.

THE EIGENVALUE PROBLEM FOR ORDI [illegible] FFERENTIAL EQUATIONS OF THE SECOND ORDER A [illegible] BERG'S THEORY OF S-MATRICES*

By KUNIHIKO KODAIRA

Recently E. C. Titchmarsh has treated the theory of expansion of arbitrary functions in terms of the eigenfunctions of a differential operator of the second order by a new method and obtained results of importance for applications.[1] The meth [illegible] Titchm [illegible] sh is based solely on the calculus of residues. In the [illegible] shall first give another proof of Titch- [illegible] arsh's results [illegible] theory of linear operators in Hilbert [illegible] ace and, secc [illegible] fferential equations of Schrödinger type [illegible] w the [illegible] concerni [illegible] the S-matrix can be fou [illegible]

1. Spect [illegible] nsider the differential expression:

$$L[u] = \text{[illegible]} dx)\{p(x)d/dx\}u + q(x)\cdot u,$$

$$(a < x < b, \; -\infty \leqq a < \text{[illegible]} \leqq +\infty)$$

defined in a (finite or infinite) open interval (a, b), where $p(x)$, $q(x)$ are real-valued functions defined in (a, b), $p(x)$ has continuous first derivative, $q(x)$ is continuous, and $p(x) > 0$ for $a < x < b$; for $x \rightarrow a$ or $x \rightarrow b$, $p(x)$, $q(x)$ may behave arbitrarily, e. g. increase infinitely, oscillate infinitely many times, etc.

Classification. Consider the differential equation

$$L[u] = l \cdot u, \tag{1.1}$$

where l means a complex parameter.

---

* Received May 12, 1949.

[1] Titchmarsh [13]. The results of the present paper were obtained by the author in August, 1948 independently of Titchmarsh's work and other recent literature which was inaccessible in Japan. The paper was revised following the suggestion of Professor H. Weyl, who kindly informed the author of the literature. The author wishes to express here his best thanks to Professor Weyl.

[2] Heisenberg [7]. See also Ma [9], Jost [8].

THEOREM 1.[illegible] *se a fixed point c, $a < c < b$, arbitrarily. If every solution u of* L[illegible] *is square summable in $(a, c]$ (or $[c, b)$) for some* l*, then, f*[illegible] *every solution u of $L[u] = l \cdot u$ is also square* [illegible]*abl*[illegible] *$[c, b)$).*[3]

Since the [illegible] u is square summable in $(a, c]$ (or $[c, b)$) or not independent[illegible] of c, we can classify, by virtue of Theorem 1.1, the differen[illegible] expressio[illegible] L in two types with respect to a (or b): if every solution u of $L[u] = l \cdot u$ is square summable in $(a, c]$ (or $[c, b)$), L is said to be of the *l. c. type* (*limit circle type*) at a (or b); otherwise L is said to be of the *l. p. type* (*limit point type*) at a (or b). Thus there exist the following four cases:

- I. L is of the l. p. type at both a a[illegible]
- II. L is of the l. c. type at a a[illegible]
- II′. L is of the l. p. type at [illegible]
- III. L is of the l. c. type at [illegible]

In what follows the case II′ will b[illegible] [illegible]duced to the case II by the transformation [illegible]

Bracket. For arbitrary functions u, v, [illegible]duce the *bracket*:

$$[uv](x) = p(x)[u(x)v'(x) - \text{[illegible]}\, u'(x)],$$

$$(u' = du/dx,\ v' = dv/dx).$$

In case u and v satisfy one and the same equation $L[u] = l \cdot u$, we write $[uv]$ for $[uv](x)$, since, in this case, $[uv](x)$ does not depend on x.

Fundamental solutions. By a *system of fundamental solutions* we shall mean the system of two solutions $s_1(x, l)$, $s_2(x, l)$ of the equation $L[u] = l \cdot u$ having the following properties:

$$(1.2)\quad \begin{cases} \text{i)} & [s_2 s_1] = 1, \\ \text{ii)} & \overline{s_k(x, l)} = s_k(x, \bar{l}), \\ \text{iii)} & \textit{as functions of } l,\ s_k(x, l) \textit{ and } (d/dx) s_k(x, l) \end{cases} \qquad (k = 1, 2),^{5}$$

($k = 1, 2$) *are regular analytic in the whole l-plane.*

---

[3] Weyl [14], Theorem 5, p. 238.

[4] Weyl [14], [15].

[5] The bar means the conjugate complex number.

Such a system of solutions s_1, s_2 is obtained, fo [illegible] ple, by solving the equation $L[u] = l \cdot u$ under the boundary conditi [illegible]

$$s_1(c) = s'_2(c) = 0, \qquad s_2(c) = p(c)s'_1(c) \quad [illegible] \quad c < [illegible]$$

where c is an arbitrary fixed point in (a, b).

The differential operator. Denote by $\mathfrak{H}$ the [illegible] consisting of all square summable functions defined in (a, b) [illegible] *duct* of the functions u, v in $\mathfrak{H}$ will be denoted by (u, v), an [illegible] u by $\| u \|$. In order to consider L as a linear operator on $\mathfrak{H}$, we ha [illegible] make explicit the *domain* of L. For that purpose, we introduce the subspace $\mathfrak{D}$ of $\mathfrak{H}$ consisting of all functions u having the following properties:

i) $u \varepsilon \mathfrak{H}$,

ii) u is [illegible] *pen* interval (a, b),

iii) du/d [illegible] in every closed subinterval $[a_1, b$ [illegible] $, b)$,

and consider L [illegible] ng $\mathfrak{D}$ as its domain. Then L becomes a *closed* operator, [illegible] noted by T; i. e. we put

$$[illegible] = L[u], \qquad \text{for } u \varepsilon \mathfrak{D}.$$

In the case I, *T is sel* [illegible] *nt; while, in the cases* II *and* III, [illegible] *s not self-adjoint.*[7]

Boundary conditions. Suppose L to be of the l. c. type at a. Then, if $u \varepsilon \mathfrak{D}$, the limit:

$$[wu](a) = \lim_{x \to a} [wu](x)$$

exists for an arbitrary function $w(x)$ defined in $(a, c]$ $(a < c < b)$ having continuous second derivatives such that

$$\int_a^c |w|^2 \, dx < +\infty, \qquad \int_a^c |L[w]|^2 \, dx < +\infty. \tag{1.3}$$

Fixing such a function w, we take therefore

$$[wu](a) = 0 \tag{1.4}$$

---

[6] Stone [11], Theorem 10. 11.

[7] Weyl [14], Chap. III and Chap. II, respectively. These statements can be found in Stone [11], Theorem 10. 12 in the terminology employed here.

as the boundary co [illegible] at a. The condition (1.4) will be called *trivial*, if every $u \varepsilon \mathfrak{D}$ s [illegible] 4). From the identity:

[illegible] $[ws_1](a)[s_2u](a) - [ws_2](a)[s_1u](a)$

s [illegible] *dary condition* (1.4) *is not trivial if and only if*

(1.6) [illegible] $](a)|^2 + |[ws_2(l)](a)|^2 \neq 0$

for fixed l, w [illegible] (l) ([illegible] $= 1, 2)$ mean the fundamental solutions mentioned above.

Now, in the case I, we put

$$H = T, \tag{1.7$_I$}$$

since T is self-adjoint. In the case II, we [illegible] d function $w_a(x)$ satisfying (1.3) and (1.6), and restr [illegible] T by imposin[g] the non-trivial boundary condition [illegible] get from T self-adjoint operator,[8] [w]hich will b[e] [illegible] put

(1.7$_{II}$) $Hu = L[u]$, for [illegible]

Finally, in the case III, we take, beside [illegible] entioned above, the real valued function $w_b(x)$ in $[c, b)$ [illegible] conditions at b corresponding to (1.3) and (1.6), and res [illegible] main $\mathfrak{D}$ of T by imposing the boundary conditions:

$$[w_au](a) = [w_bu](b) = 0.$$

Then we obtain from T a self-adjoint operator,[9] which will be denoted by H; i. e. we put

$$Hu = L[u], \text{ for } u \varepsilon \mathfrak{D},\ [w_au](a) = [w_bu](b) = 0. \tag{1.7$_{III}$}$$

Our purpose is to determine the *spectra of the self-adjoint operator* H thus defined.

Characteristic functions. Let a self-adjoint operator H defined as above be given. Then we put

$$f_b(l) = -\lim_{x \to b} s_2(x, l)/s_1(x, l) \tag{1.8$_b$}$$

if L is of the l. p. type at b, and

---

[8] Weyl [14], Chap. II. Cf. also Stone [11], Theorem 10.13.

[9] Weyl [15], pp. 241-242. Cf. also Stone [11], Theorem 10.14.

(1.9_b) $f_b(l) = -\{[w_b s_2(l)](b)\}/\{[w_b s$

if L is of the l. c. type at b; similarly we put

(1.8_a) $f_a(l) = -\lim_{x \to a} s_2(x, l)/s_1(x, l$

if L is of the l. p. type at a, and

(1.9_a) $f_a(l) = -\{[w_a s_2(l)](a)\}/\{[w$

if L is of the l. c. type at a. The functions $f_b(l)$ a thus defined will be called the *characteristic functions*[10] of the operator H. In case L is of the l. p. type at b (or a), the function $f_b(l)$ (or $f_a(l)$) is uniquely determined by the condition[11]

(1.10_b) $\int_c^b |$ $(x, l)|^2 dx < +\infty,$ $(a < c < b)$

(1. $)|^2 dx <$, $(a < c < b).$

From ,

(1.11) $f_b(\bar{l}) = \overline{f_b(l)}.$

Again we have

(1.12) $\neq f_b(l),$ (f $\neq 0$),

since H is self-adjoint.

THEOREM 1.2. *The characteristic functions $f_a(l)$ and $f_b(l)$ are analytic functions of l which are meromorphic in $\Im l \neq 0$. Especially, if L is of the l. c. type at a (or b), then $f_a(l)$ (or $f_b(l)$) is meromorphic also on the real axis.*[12]

Spectral theorem. Let H be a self-adjoint differential operator defined as above and

$$H = \int_{-\infty}^{+\infty} \lambda dE(\lambda)$$

---

[10] Titchmarsh, Chap. III. Our theory is parallel to that of Jacobi matrices. See, e. g., Stone [11], Chap. X, § 4. The characteristic functions correspond to the function $F(l)$ in Stone [11], p. 560.

[11] Weyl [14], pp. 227. Cf. also Stone [11], Theorems 10.12, 10.14, 10.20.

[12] For the proof of this theorem, see Weyl [16].

on of H. Introduce the "*characteristic matrix*" by

$$) = f_a(l) \cdot f_b(l) \cdot [f_a(l) - f_b(l)]^{-1},$$
$$M_{21}(l) = \tfrac{1}{2} \cdot [f_a(l) + f_b(l)] \cdot [f_a(l) - f_b(l)]^{-1},$$
$$_{22}(l) = [f_a(l) - f_b(l)]^{-1},$$

where $f_a(l)$, haracteristic functions of H. It follows from (1.11), (1. em 1.2 that $M_{jk}(l)$ is an analytic function of l which is regu $\neq 0$ and

$$M_{jk}(\bar{l}) = \overline{M_{jk}(l)}. \tag{1.14}$$

THEOREM 1.13. (SPECTRAL THEOREM).[13] *For every real number λ there exists the limit*

$$\rho_{jk}(\lambda) = \lim_{\delta \to +0} \lim_{\epsilon \to +0} 1/\pi \int \tag{1.15}$$

As a function of λ, t matrix fun conti on the right and monoto non-decrea he symmetric matrix $\mathrm{P}(\lambda) - \mathrm{P}(\mu)$ is p ery finite interval $\Delta = (\mu, \lambda]$,

$$E(\Delta) = E(\lambda) - E$$

Then, for every $u \varepsilon \mathfrak{H}$, $E(\Delta)u(x)$ can be repre lows:

$$E(\Delta)u(x) = \int_a^b u(y)\,dy \int_\Delta \sum_{j,k} s_j(x,\lambda) s_k(y,\lambda)\, d\rho_{jk}(\lambda), \tag{1.16}$$

where

$$\int_a^b dy \left| \int_\Delta \sum_{j,k} s_j(x,\lambda) s_k(y,\lambda)\, d\rho_{jk}(\lambda) \right|^2 < +\infty$$

and the integral $\int_a^b dy$ in (1.16) converges absolutely. $u(x)$ itself is repre sented as follows:

$$u(x) = \lim_{\substack{\lambda \to +\infty \\ \mu \to -\infty}} \int_a^b u(y)\,dy \int_\mu^\lambda \sum_{j,k} s_j(x,\lambda) s_k(y,\lambda)\, d\rho_{jk}(\lambda), \tag{1.17}$$

where the limit converges in the mean; especially if u belongs to the doma of H, we have

$$Hu(x) = \lim_{\substack{\lambda \to +\infty \\ \mu \to -\infty}} \int_a^b u(y)\,dy \int_\mu^\lambda \sum_{j,k} s_j(x,\lambda) s_k(y,\lambda) \lambda\, d\rho_{jk}(\lambda). \tag{1.18}$$

---

[13] Cf. Titchmarsh [13], Chap. III.

2. Proof of the spectral theore

heorem in the special case that the
onditions

$$(2.1) \qquad s_1(c) = s'_2(c) = 0, \qquad s_2($$

For every l with $\Im l \neq 0$, we introd
lefined by

$$G(x, y, l) = G(y, x, l) = g_a(x, l) g$$

here

$x, \Delta)$. For every $u \varepsilon \mathfrak{H}$, $u(\Delta)$ belongs to the domain   or

$$v(\Delta) = (H - l) u(\Delta) = \int_\Delta (\lambda - l)\, dE(\lambda) u,$$

have therefore, by (2. 2), $u(\Delta) = G(l) v(\Delta)$. Hence we have.

$$u(x, \Delta) = \int_a^b G(x, y, l) v(y, \Delta)\, dy = (v(\Delta), \overline{G(x, \ , l)})$$
$$= (\int_\Delta (\lambda - l)\, dE(\lambda) u, \overline{G(x, \ , l)}) = (u, \int_\Delta (\lambda - \bar{l})\, dE(\lambda) \overline{G(x, \ , l)}).$$

us we obtain the formulae

---

[17] We follow the method of H. Weyl. Cf. Weyl [14], pp. 238-251.

[18] Weyl [14], pp. 224-231. Cf. also Stone [11], Theorem 10. 19, 10. 20 and 10. 21.

$$= (u, \int_\Delta (\lambda - \bar{l})\, dE(\lambda)\, \overline{G(x,\ ,l)}),$$

$$u, \int_\Delta (\lambda - \bar{l})\, dE(\lambda)\, \overline{G_x(x,\ ,l)}), \quad (G_x = dG/dx).$$

follows that the functions $u(x, \Delta)$, $(d/dx)u(x, \Delta)$ are continuous w x and of bounded variation with respect to λ in every finite Considering $u(\Delta)$ as an element of $\mathfrak{H}$, we have the relation

$$L[u(\Delta)] = Hu(\Delta) = \int_\Delta \lambda\, dE(\lambda)\, u(\Delta),$$

showing that the function $u(x, \lambda)$ satisfies the integro-differential equation:

$$L[u(x, \lambda)] = \int_0^\lambda \lambda d \qquad (2.5)$$

$u(x, \lambda)$ satisfies furthermore the boun

$$(x, 0) = (d/d$$

Under these circumstances, the solutio) is given by

$$u(x, \lambda) = \int_0^\lambda \{s_1(x, \lambda)\, du_1(\lambda) + \quad)\}, \qquad (2.6)$$

where

$$(\lambda) = p(c)u'(c, \lambda), \qquad u_2(\lambda) = u(c, \lambda), \qquad (u' = du/dx).^{19} \qquad (2.7)$$

In what follows we assume that the additive functions $u_k(\Delta)$, $\xi_k(\Delta)$, $\rho_{jk}(\Delta)$, etc. of the interval Δ are always combined with the corresponding functions $u_k(\lambda)$, $\xi_k(\lambda)$, $\rho_{jk}(\lambda)$, etc. by the relations as follows:

$$u_k(\lambda) = u_k((0, \lambda]); \qquad u_k(\Delta) = u_k(\lambda) - u_k(\mu), \qquad (\Delta = (\mu, \lambda]).$$

Now, putting $\gamma_1(l) = p(c)\overline{G_x(c, l)}$, $\quad \gamma_2(l) = \overline{G(c, l)}$ and

$$\xi_k(\Delta) = \int_\Delta (\lambda - \bar{l})\, dE(\lambda)\, \gamma_k(l), \qquad (k = 1, 2), \qquad (2.8)$$

we get from (2.3), (2.4) and (2.7)

$$u_k(\Delta) = (u, \xi_k(\Delta)), \qquad (k = 1, 2). \qquad (2.9)$$

For fixed Δ, $u_k(\Delta)$ can be considered as linear functionals of u, not depending

---

[19] Weyl [14], Hilfssatz 1 and its proof, pp. 240-241.

on l. Hence we infer, by (2.9), that the functions

Insert now $u=\xi_j(\Delta)$, $(\Delta=(0,\lambda])$ in (2.6). T

$$\xi_j(x,\Delta)=\int_\Delta \{s_1(x,\lambda)\,d\rho_{j1}(\lambda)+s_2(x,\lambda \tag{2.10}$$

where

$$\rho_{jk}(\Delta)=(\xi_j(\Delta),\xi_k(\Delta \tag{2.11}$$

Inserting (2.8) in (2.11), we get

$$\rho_{jk}(\Delta)=\int_\Delta |\lambda-l|^2(dE(\lambda)\gamma_j(l),\gamma_k(l)),$$

whence we conclude

$$\int^{+} \quad l|^{-2}d\rho_{jk}(\lambda)=(\gamma_j(l),\gamma_k(l)), \tag{2.12}$$

here the integ utely. From (2.11) follows that the
mat $(\Delta)=$ e semi-definite d independent on l.
Us $y, l)$, we can r dily calculate the right
han we obtain

$$j,\gamma_k(l))=\Im M_{jk}(l),$$

and therefore th

$$l|^{-2}\Im l\,d\rho_{jk}(\lambda)=\Im M_{jk}(l), \tag{2.13}$$

which yields immediately the spectral formula (1.15). To prove (1.16), we choose real numbers $\lambda_0,\lambda_1,\cdots,\lambda_n$ so that

$$\mu=\lambda_0<\lambda_1<\cdots<\lambda_n=\lambda,$$

put $\delta=\max\limits_m|\lambda_m-\lambda_{m-1}|$, and consider the sum

$$S_\delta(y)=\sum_m\sum_j s_j(x,\lambda_m)\xi_j(y,\Delta_m), \qquad (\Delta_m=(\lambda_{m-1},\lambda_m]).$$

Then we have, by virtue of (2.10),

$$\lim_{\delta\to 0}S_\delta(y)=\int_\Delta \sum_{j,k}s_j(x,\lambda)s_k(y,\lambda)\,d\rho_{jk}(\lambda),$$

while, since $\sum\limits_m\|\xi_j(\Delta_m)\|^2=\|\xi(\Delta)\|^2$, there exists $\lim S_\delta$ in the sense of $\|\ \ \|$ and thus $\lim S_\delta\varepsilon\mathfrak{H}$. This proves the inequality

$$\int_a^b dy\,\Big|\int_\Delta \sum_{j,k}s_j(x,\lambda)s_k(y,\lambda)\,d\rho_{jk}(\lambda)\Big|^2<+\infty.$$

and (2. 9),

$$_j(x, \lambda_m) u(x, \Delta_m)$$

$$\lim_{\delta\to 0} S_\delta) = (u, \int_{\Delta} \sum_{j,k} s_j(x, \lambda) s_k(\ , \lambda)\, d\rho_{jk}(\lambda)),$$

proving (1. 1 and (1. 18) follow immediately from (1. 16). Again, (1. (2. 13). Thus the spectral theorem has been proved for mentioned above.

To prove ectral theorem in general cases, we denote the special fundamental solutions s_1, s_2 and the corresponding M_{jk}, ρ_{jk} mentioned above by s_1^0, s_2^0, M_{jk}^0, ρ_{jk}^0, respectively. Then general s_1, s_2 are related to s_1^0, s_2^0 by a unimodular transformation:

$$(2.14) \qquad s_j^0(x, l) = \sum_k \beta_{jk}(l) s_k(x, l), \qquad \det(\beta_{jk}(l)) = 1,$$

where $\beta_{jk}(l)$ are holomorphic functions $= \overline{\beta_{jk}(l)}$. By tl transformation (2. 1 the characteris formed ding to the rule:

$$M_{jk}(l) = \sum_{m,n} \beta_{mj}(l)$$

This shows, combined with the relation:

$$) = \int_{-\infty}^{+\infty} \{(\lambda - l)^{-1} - (\lambda - l_0 \qquad + \text{const.},$$

that

$$(2.15) \qquad M_{jk}(l) - \int_{-\nu}^{\nu} \sum_{m,n} (\lambda - l)^{-1} \beta_{mj}(\lambda) \beta_{nk}(\lambda)\, d\rho^0_{mn}(\lambda)$$

are regular except for real l such that $l \leqq -\nu$ or $l \geqq \nu$; hence the functions $\rho_{jk}(\lambda)$ defined by (1. 15) are given by

$$(2.16) \qquad \rho_{jk}(\lambda) = \int_0^{\lambda} \sum_{m,n} \beta_{mj}(\lambda) \beta_{nk}(\lambda)\, d\rho^0_{mn}(\lambda).$$

Inserting this in (2. 15), we see immediately that $R_{jk}^{(\nu)}(l)$ are regular except for real l such that $l \leqq -\nu$ or $l \geqq \nu$. Again, from (2. 14) and (2. 16) follows that the relations (1. 16), (1. 17) and (1. 18) are preserved by the transformation $s_k^0 \to s_k$, $\rho_{jk}^0 \to \rho_{jk}$. Thus the spectral theorem is completely proved.

3. Hill's equation.

As an example, we consider

$$L = -d^2/dx^2 + q(x), \qquad (q(x+1) = q(x), -\infty < x < +\infty),$$

$s_1(0) = s'_2(0) = 0$, $s_2(0) = s'_1(0) = 1$. Then w

$$s_j(x+1, l) = \sum_{k=1}^{2} h_{jk}(l) s_k(x, l),$$

. Let $z_\pm(l)$

where $h_{jk}(l)$ are holomorphic functions of l an
be the roots of the secular equation

$$| z \cdot \delta_{jk} - h_{jk}(l) | \equiv z^2 - 2\tau(l) \cdot z + 1 = 0, \quad (2\tau(l) = h_{11}(l) + h_{22}(l))$$

such that $| z_+(l) | \geqq 1$, $| z_-(l) | \leqq 1$, and $f_\pm(l)$ be the corresponding solutions of the linear equations:

$$h_{11}(l \qquad z \cdot f, \quad h_{12}(l) \cdot f + h_{22}(l) = z.$$

The ting

(3 $\qquad l) + f_\pm(l) \cdot s_1(x, l),$

we obtain the s $\qquad L[u] = l \cdot u$ having the following properties:

(3.2) $\qquad 1, l) = z_\pm(l) \cdot g_\pm(x, l).$

Since, in general, | $\qquad$ 1, $| z_-(l) | < 1$, we infer, by (3. $\qquad$ d (3.1), that our L belongs to the case I and $f_{\pm\infty}(l) = f_\mp(l)$. Hence we get

$$(3.3) \qquad M(l) = 4^{-1}\{\tau^2(l) - 1\}^{-\frac{1}{2}} \begin{pmatrix} 2h_{21}(l) & h_{22}(l) - h_{11}(l) \\ h_{22}(l) - h_{11}(l) & -2h_{12}(l) \end{pmatrix}.$$

$\Im M_{jk}(\lambda + i \cdot 0)$ is therefore $\neq 0$ if and only if $\tau^2(\lambda) - 1 < 0$, since $h_{jk}(\lambda)$ and $\tau(\lambda)$ are real for real λ. Now, it is known[21] that the equation $\tau^2(\lambda) - 1 = 0$ has infinitely many roots $\lambda_0, \lambda_1, \lambda_2, \cdots$ such that

$$\lambda_0 < \lambda_1 \leqq \lambda_2 < \lambda_3 \leqq \lambda_4 < \cdots \leqq \lambda_{2m} < \lambda_{2m+1} \leqq \cdots, \lambda_m \to +\infty \ (m \to \infty)$$

and that

---

[20] Many examples of applications to the spectral theorem are to be found in Titchmarsh [13]. Here we shall consider Hill's equation, which has been treated by Wintner [20], [18]. Wintner has determined the spectrum of this equation; above, the matrix $P(\lambda)$ determining the spectral resolution and, therefore, the spectrum, will be obtained.

[21] Cf. Strutt [12].

$$
\begin{matrix}
0, & \text{for } \lambda < \lambda_0 \text{ and } \lambda_{2m-1} < \lambda < \lambda_{2m}, \\
0, & \text{for } \lambda_{2m} < \lambda < \lambda_{2m+1}.
\end{matrix}
$$

ula (1. 15), we get therefore, from (3. 3)

$$
dP(\lambda) = \begin{cases} 1 - \tau^2(\lambda))\}^{-\frac{1}{2}} \begin{pmatrix} -2h_{21}(\lambda), & h_{11}(\lambda) - h_{22}(\lambda) \\ h_{11}(\lambda) - h_{22}(\lambda), & 2h_{12}(\lambda) \end{pmatrix} d\lambda, \\ \qquad\qquad\qquad\qquad\qquad\qquad (\text{for } \lambda_{2m} \leqq \lambda \leqq \lambda_{2m+1}), \\ \qquad\qquad\qquad\qquad\qquad\qquad (\text{otherwise}). \end{cases}
$$

Thus we obtain, by (1. 23),

$$
\begin{aligned}
u(x) = \sum_{m=0}^{\infty} \int_{\lambda_{2m}}^{\lambda_{2m+1}} & (-1)^m (4\pi)^{-1} \{1 - \tau^2(\lambda)\}^{-\frac{1}{2}} \\
& \times \int_{-\infty}^{+\infty} u(y) \{2h_{12}(\lambda) s_2(x, \lambda) s_2(\\
- h_{21}(\lambda) s_1(x, \lambda) & (y, \lambda) + [h_{11}(\lambda) - \qquad\qquad (y, \lambda) \\
& + s_1(x, \quad) s_2
\end{aligned}
$$

The operator $H = -d^2/dx^2 + q(x)$ has th pectrum and the continuous spectrum of H consists of infinite rvals $[\lambda_{2m}, \lambda_{2m+1}]$ $(m = 0, 1, 2, \cdots)$.

4. **case that $f_a(l)$ is a meromorphic** As one readily verifies, $f_a(l)$ is a meromorphic function or not, independently of the choice of the system of fundamental solutions s_1, s_2, and, in case L is of the l. c. type at a, of the boundary condition. *In case $f_a(l)$ is a meromorphic function, the system of fundamental solutions s_1, s_2 can be chosen so that $f_a(l) = \infty$ identically in l.* Such a system of fundamental solutions will be called *normal.* In case L is of the l. c. type at a, we have $f_a(l) = \infty$ if and only if $s_1(l)$ satisfies $[w_a s_1(l)](a) = 0$; in case L is of the l. p. type at a, $f_a(l) = \infty$ if and only if

$$
\int_a^c |s_1(x, l)|^2 dx < \infty, \qquad (a < c < b).
$$

Now, in case $f_a(l) = \infty$ identically in l, the characteristic matrix $M(l)$ has the form

$$
M(l) = \begin{pmatrix} f_b(l) & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}; \text{ consequently, } P(\lambda) = \begin{pmatrix} \rho(\lambda) & 0 \\ 0 & 0 \end{pmatrix}.
$$

Thus, in this case, the spectral theorem becomes as follows:

THEOREM 4.1. (SPECIAL FORM OF THE SPEC[illegible] *that* $f_a(l) = \infty$ *identically in* l. *Then* $f_b(l)$ *is re*[illegible] *For every real number* λ *there exists the limit:*

$$\rho(\lambda) = \lim_{\delta \to +0} \lim_{\epsilon \to +0} 1/\pi \int_{\delta}^{\lambda+\delta} \Im f_b(\lambda + i\epsilon) [illegible] \tag{4.1}$$

$\rho(\lambda)$ *is a monotone non-decreasing function of* λ. [illegible]*ous on the right. Let* $H = \int_{-\infty}^{+\infty} \lambda dE(\lambda)$ *be the spectral representa*[illegible] *H, and put* $E(\Delta) = E(\lambda) - E(\mu)$ *for every finite interval* $\Delta = (\mu, \lambda]$. *Then, for arbitrary* $u \in \mathfrak{H}$, *we have*

$$E(\Delta)u[illegible] {}^{b} u(y)\,dy \int_{\Delta} s_1(x,\lambda)s_1(y,\lambda)\,d\rho(\lambda), \tag{4.2}$$

[illegible]*ere*

$$\int [illegible] (y,\lambda)\,d\rho(\lambda)|^2 [illegible] \infty$$

and [illegible] *egra*[illegible] *nverges absolutely.* $u(x)$ *itself is represented as follows:*

$$u(x [illegible] u(y)\,dy \int_{\mu}^{\lambda} s_1(x,\lambda)s_1(y,\lambda)\,d\rho [illegible] \tag{4.3}$$

where the limit converges in the mean. If, especially, u *be*[illegible]*gs to the domain of* H, *we have*

$$Hu(x) = \lim_{\substack{\lambda \to +\infty \\ \mu \to -\infty}} \int_a^b u(y)\,dy \int_{\mu}^{\lambda} s_1(x,\lambda)s_1(y,\lambda)\lambda\,d\rho(\lambda). \tag{4.4}$$

$\rho(\lambda)$ *is represented also in the following form:*

$$\rho(\lambda) = -\lim_{\mu \to \lambda+0} \lim_{\epsilon \to +0} (2\pi i)^{-1} \int_{C(\mu,\alpha,\epsilon)} f_b(l)\,dl, \tag{4.5}$$

where $C(\mu, \alpha, \epsilon)$ *means the contour consisting of two oriented polygonal lines whose vertices, in order, are* $\mu + i\epsilon$, $\mu + i\alpha$, $i\alpha$, $i\epsilon$, *and* $-i\epsilon$, $-i\alpha$, $\mu - i\alpha$, $\mu - i\epsilon$, *respectively, the real number* α *being subject to the inequality* $\alpha > \epsilon$. *For every* $\nu > 0$, *the residual term*

$$r^{(\nu)}(l) = f_b(l) - \int_{-\nu}^{\nu} (\lambda - l)^{-1} d\rho(\lambda)$$

xcept for real l such that $l \leqq -\nu$ or $l \geqq \nu$. Thus in the finite l-plane are completely determined by isolated singular point of $f_b(l)$ is a pole of the order 1.

mulae (1.24), (1.25), etc., and Theorem 1.4 are dingly s as follows: Consider λ-measurable functions $\phi(\lambda)$, define the n

$$\phi \|^2 = \int_{-\infty}^{+\infty} | \phi(\lambda) |^2 \, d\rho(\lambda),$$

and introduce the Hilbert space $\mathfrak{H}^* = \{\phi \mid \| \phi \| < +\infty\}$. Then we have

THEOREM 4.2. *The transformation*

$$u(x) \rightarrow \phi(\lambda) = \int_a^b s_1(\tag{4.6}$$

is a unitary transfor ion mapping $\mathfrak{H}$ *erse is by*

$$\phi(\lambda) \rightarrow u(x) = \int_{-\infty}^{+} s \tag{4.7}$$

where the integral in (4.6) *or* (4.7) *conver nse of* $\| \quad \|$ *in* $\mathfrak{H}$ *or* $\mathfrak{H}^*$, *respe ely.*

By e of (4.6) and (4.7), the form can be rewritten as follows:

$$u(x) = \int_{-\infty}^{+\infty} s_1(x, \lambda) \, d\rho(\lambda) \int_a^b s_1(y, \lambda) u(y) \, dy. \tag{4.8}$$

Again, (4.4) shows that the transformation $u \rightarrow \phi$ transforms H into the "diagonal form" $\lambda \times$.

In the case that $f_a(l)$ and $f_b(l)$ are both meromorphic functions, the structure of H is a very simple one. To see this, we choose s_1, s_2 so that $f_a(l) = \infty$ identically. The corresponding $f_b(b)$ is then a meromorphic function which is regular in $\Im l \neq 0$. Denote the poles of $f_b(l)$ by λ_m $(m = 1, 2, 3, \cdots)$ and put

$$\rho_m = -(2\pi i)^{-1} \oint_{\lambda_m} f_b(l) \, dl.$$

Then we have, by (4.5),

$$\rho(\lambda) - \rho(\mu) = \sum_{\mu < \lambda_m \leqq \lambda} \rho_m, \tag{4.9}$$

while, since the poles λ_m are all of the order 1, ρ_m becomes therefore as follows:

$$u(x) = \sum_m s_1(x, \lambda_m)\rho_m \int_a^b s_1(y, \lambda_m) u \tag{4.10}$$

Thus, in this case, H has no continuous spectrum and of H is a discrete set consisting of $\lambda_1, \lambda_2, \lambda_3, \cdots$ if H has no continuous spectrum and the point spectrum of , $f_a(l)$ and $f_b(l)$ are both meromorphic functions, since, by are meromorphic functions. We conclude:

THEOREM 4. 3.[22] *H has only the discrete point spectrum if and only if $f_a(l)$ and $f_b(l)$ are both meromorphic functions.*

What is the conditio $f_a(l)$ to be a meromorphic function? *In case is of the l. c. t is always a meromorphic function,* as was ready mentioned ; while, in case L is of the l. p. type at a, the s u is con e, for simplicity' ke, that L is analytic in and $q(x)$ are alytic functions in a neig for example,

THEOREM 4. *that $a > -\infty$ and a is the regular singular point of the diffe on $L[u] = 0$. Then, $f_a(l)$ is a meromorphic function, if*

$$)^2/p(x) \rightarrow 0 \qquad \rightarrow a).$$

Combined with Theorem 4. 3, this yields immediately the following

THEOREM 4. 5. *Assume that L is analytic, $-\infty < a$, $b < +\infty$ and a, b are both regular singular points of the differential equation $L[u] = 0$. Then H has only the discrete point spectrum, if*

$$(x-a)^2/p(x) \rightarrow 0 \quad (x \rightarrow a), \qquad (x-b)^2/p(x) \rightarrow 0 \quad (x \rightarrow b).$$

5. Schrödinger equation. Let us consider the differential operator of *Schrödinger type*:

$$L = -d^2/dx^2 + \nu(\nu+1)/x^2 + V(x), \qquad (0 < x < \infty),$$

where ν is a real number $\geqq -\frac{1}{2}$ and $V(x)$ is a real continuous function such that

---

[22] Cf. Titchmarsh [13], Chap. II.

$), \qquad V(x) = [\alpha + O(x^{-\epsilon})]/x$, for $x \to \infty$, $(\epsilon > 0)$.

damental solutions s_1, s_2 of $L[u] = lu$ (satisfying so that, as $x \to 0$,

$$x^{\nu+1}; \qquad s_2(x, l) \sim (2\nu + 1)^{-1}/x^{\nu}, \nu > -\tfrac{1}{2};$$

$$) \sim -x^{\frac{1}{2}} \log x, \nu = -\tfrac{1}{2}.$$

We conclude hat, *at* 0, *L is of the l. p. type, if* $\nu \geqq \frac{1}{2}$, *and of the l. c. type, if* $\nu < \frac{1}{2}$. In case $\nu \geqq \frac{1}{2}$, the system of fundamental solutions s_1, s_2 defined above is *normal* (i. e. $f_a(l) = \infty$), since

$$\int_0^1 | s_1(x, l) |^2 \, dx < +$$

In case $\nu < \frac{1}{2}$, we take, as the boundary con

(5.1) $\qquad [w \quad (0) = 0, \qquad w$

then the system s_1, s_2 is also *normal*. Th 4.2 can be applied here.

In order to investigate the asymptotic the solution of $L[u] = l \cdot u$ o, we put

$$k = l^{\frac{1}{2}}, \qquad l = k.$$

Then we have

THEOREM 5.1. *If* $\Im k \geqq 0$, $k \neq 0$, *the equation* $L[u] = k^2 u$ *has one and only one solution* $u(x, k)$ *such that*

(5.2) $$u(x, k) \sim \exp[ikx - \tfrac{1}{2}i\alpha/k \log x], \qquad \text{for } x \to \infty.$$

As functions of two variables x, k, $u(x, k)$ *and* $u'(x, k)$ *are continuous in* $0 < x < \infty$, $\Im k \geqq 0$, $k \neq 0$; *as functions of* k, $u(x, k)$ *and* $u'(x, k)$ *are regular analytic in* $\Im k > 0$. *Furthermore we have*

(5.3) $$u'(x, k) \sim (d/dx) \exp[ikx - (\tfrac{1}{2}i\alpha/k) \log x], \qquad \text{for } x \to \infty.$$

Proof. Put $u = v \cdot \exp[ikx - (\frac{1}{2}i\alpha/k) \log x]$. Then the equation $L[u] = k^2 u$ is reduced to

(A) $$[p(x, k) \cdot v']' - p(x, k) V^*(x, k) \cdot v = 0,$$

where

$$\begin{cases} p(x,k) = \exp[2kix - (i\alpha/k)\log x], \\ V^*(x,k) = [\alpha/2ik + \alpha^2/4k^2]/x^2 + [V(x) \end{cases}$$

thus we have first to show that the differential equatio only one solution $v = v(x,k)$ satisfying $v(x,k) \to 1$ This ca readily done by transforming the differential e the integral equation

$$\text{(B)} \qquad v(x) = g(x,k) - \int_x^\infty K(x,y,k)v(y$$

where $g(x,k) = (1-\alpha/2k^2x)^{-1}$ and

$$K(x,y,k) = (p(x,k)/p' \quad k))\{[1-p(y,k)/p(x,k)]V^*(y,k)$$
$$) (p'(y,k)/p(y,k))\}, \qquad (x \leqq y).$$

urthermore, in t way, solving the integral equation (B) by the of ite readily that the ution $v(x,k)$ of (B) wit derivative $v'(x,k)$ are continuous in $0 <$ k at, as functions of k, $v(x,k)$ and $v'(x,k)$ are holomorphic Thus Theorem 5.1 is proved.[23]

From (5.2)

$$\text{(5.4)} \qquad \int_1 \quad dx \leqq +\infty \text{ according as } \Im k \geqq$$

whence we conclude that L is of the l. p. type at ∞. We need therefore no boundary condition at ∞.

Put now

$$u(x,k) = A(k)s_2(x,l) - B(k)s_1(x,l), \qquad (k^2 = l, \Im k \geqq 0).$$

Then we have, by (1.10_b) and (5.4),

$$\text{(5.5)} \qquad f_\infty(l) = -B(k)/A(k), \qquad (k^2 = l, \Im k \geqq 0).$$

$A(k)$ and $B(k)$ are given by

$$\text{(5.6)} \qquad A(k) = [u(k)s_1(k^2)], \quad B(k) = [u(k)s_2(k^2)].$$

Hence, by Theorem 5.1, $A(k)$, $B(k)$ are regular in $\Im k > 0$ and continuous since $u(x, k)$ is not identically 0, $A(k)$ and $B(k)$ oint. We conclude therefore, by virtue of (5. 5), phic except for $l \geqq 0$ and every pole of $f_\infty(l)$ which ids one-to-one to the zero point of $A(k)$, since, by the on h $\Im k > 0$ corresponds one-to-one to l which is not $\geqq 0$. On the other em 4. 1, every pole of $f_\infty(l)$ lies on the real axis and is of th e, in $\Im k > 0$, all zero points of $A(k)$ lie on the imaginary of the order 1. Denote these zero points by $k_m = i\,|\,k_m\,|$, 1, 2, 3, $\cdots$); then the poles of $f_\infty(l)$ which are not $\geqq 0$ are given by $\lambda_m = k^2{}_m = -\,|\,k_m\,|^2$, $(m = 1, 2, 3, \cdots)$. Now, for real $k \neq 0$, we conclude from (5. 2) and (5. 3) the formulae

---

[23] After the variation of constants employed in the proof above, the theorem 5.1 is contained in a result of Bôcher [2]; cf. also Wintner [17], [19], [21].

(5. 7) $\quad u(x, -k) = \overline{u(x, k)}, \ (k > 0)\,;$ $\quad [u($ $u(k)] = 2ik, \ (k > 0)\,;$

which yields

(5. 8) $\quad A(-k)$ $\overline{A(k)}, \quad B(-k$ $),$

(5. 9) $\quad A(k)B(-k) - A(-k)B(k$ $).$

Hence $A(k)$ and $B(k)$ do not vanish for real k

Nov these results, we can readily function $\rho(\lambda)$. First, fo 0, we have

$$\rho(\lambda) = \frac{1}{\pi}\int_0^\lambda \Im f_\infty(\lambda + i0)\,d\lambda = -\frac{1}{\pi}\int_0^\lambda \Im\{B(\lambda^{\frac{1}{2}})/A(\lambda^{\frac{1}{2}})\}\,d\lambda,$$

while, by (5. 8) and (5. 9),

$$\Im\{B(k)/A(k)\} = -k/|\,A(k)\,|^2, \qquad (k > 0).$$

Hence we obtain

$$d\rho(\lambda) = (2/\pi)\,(k^2/|\,A(k)\,|^2)\,dk, \qquad (k = \lambda^{\frac{1}{2}}, \lambda > 0).$$

Secondly, for $\lambda < 0$, $\rho(-0) - \rho(\lambda)$ is, by virtue of (4. 5), equal to the sum of the residues ρ_m of $-f_\infty(l)$ at the poles λ_m, $\lambda < \lambda_m < 0$, i. e.,

$$(5.10) \quad \rho(-0) - \rho(\lambda) = \sum_{\lambda < \lambda_m < 0} \rho_m, \text{ where } \rho_m = \frac{1}{\pi}\,|\,k_m\,|\,B(k_m) \oint_{k_m} dk/A(k).$$

Obviously ρ_m is positive. Finally, putting $\rho_0 = \rho(0) - \rho(-0)$, we have

$$\rho_0 = -(1/2\pi i) \lim_{\epsilon\to 0} \oint_{|l|=\epsilon} f_\infty($$

or

$$\rho_0 = \lim_{\epsilon\to +0} \tfrac{1}{\pi} \int_0^\pi \{B(\epsilon e^{i\theta})/A(\epsilon e^{i\theta})\}\epsilon^2 e^{2i\theta} d \tag{5. 11}$$

ρ_0 is $\geqq 0$. Thus, in this case, the formula (4. ows:

$$u(x) = \sum_{m=0}^{\infty} s_1(x, \lambda_m)\rho_m \int_0^\infty s_1(y, \lambda_m)u(y)\,dy$$
$$+ (2/\pi) \int_0^\infty s_1(x, k^2)k^2/|A(k)|^2\,dk \int_0^\infty s_1(y, k^2)u(y)\,dy,$$

where $\lambda_0 = 0$. We conc therefore

THEOREM 5. 2. *int spectrum of the Schrödinger operator* $-H = -$ $/dx^2 + v$ $V(x)$ *consists of all points* $\lambda_m = k^2_m$ $= -$ m *nd, eventually, of e point* $\lambda_0 = 0$. *The cont nterval* $[0, \infty)$, *if* 0 *does not belong to the point ctrum, he other case. The corresponding "normalized eigenfunctions"*

$$u_m(x) = \rho_m^{\frac{1}{2}} \quad = 0, 1, 2, \cdots;$$
$$u_k(x) = \quad /A(k)|\, s_1(x, \lambda), \quad k = \lambda^{\frac{1}{2}} > 0,$$

where ρ_m *are the quantities defined by* (5. 10), (5. 11), *and* $\rho_m > 0$ $(m = 1, 2, \cdots)$, $\rho_0 \geqq 0$. *By means of these eigenfunctions, every square summable function* $u(x)$ *can be expanded in the following form*:

$$u(x) = \sum_{m=0}^{\infty} u_m(x) \int_0^\infty u_m(y)u(y)\,dy + \int_0^\infty u_k(x)\,dk \int_0^\infty u_k(y)u(y)\,dy.$$

For $k > 0$, we conclude, using (5. 7),

$$s_1(x, k^2) = 1/2ik\{A(-k)u(x, k) - A(k)u(x, -k)\}, \quad (k > 0). \tag{5. 12}$$

Putting

$$A(-k) = \overline{A(k)} = |A(k)|\, e^{i\delta(k)}, \qquad (k > 0), \tag{5. 13}$$

---

[24] Cf. Titchmarsh [12], Chap. V. The location of the continuous spectrum can be read off from more general theorems; see Wintner [18]; Hartman and Wintner [4], [5], [6]; Hartman [3].

$$\pi)^{\frac{1}{2}}(2i)^{-1}\{e^{i\delta(k)}u(x,k) - e^{-i\delta(k)}u(x,-k)\}.$$

ed with (5.2), the *asymptotic formula*:

$$(5.14) \qquad u \qquad \text{n}\,[kx - \alpha/2k \log x + \delta(k)], \qquad (x \to \infty).$$

6. Hei -function. In this section we shall investigate the Schrödinger operator treated in **5** under the following assumption:

Assumption I. *The functions $u(x,k)$ and $u'(x,k)$ introduced in Theorem 5.1 can be extended analytically over whole lower half k-plane across the negative part of the real axis.*

Under this assumption, the functions A) are also extende analytically over the le lower half plan egative f the real axis, and the re ions (5.8), (5. are extended for $\Im k > 0$, i. e., we have

$$(6.1) \quad A(-\bar{k}) = \overline{A(k)}, \qquad B(-\bar{k}) = \overline{B(k)} \qquad (\Im k > 0),$$

$$(6.2) \quad -k) - A(-k)B(k) = 2ik, \qquad (\Im k > 0),$$

$$(6.3) \quad s_1 \quad k^2) = (2ik)^{-1}\{A(-k)u(x,k) - A \quad -k)\}, \qquad (\Im k > 0).$$

Now we make furthermore

Assumption II. *$u(x,k)$ and $u'(x,k)$ are regular in the whole k-plane except for $k \geqq 0$.*

Then the functions $A(k)$ and $B(k)$ are also regular in the whole k-plane except for $k \geqq 0$. Furthermore, using (6.1), (6.2) and (6.3), we can eliminate the function $B(k)$ from the formula (5.10). In fact, inserting $k = k_m$ in (6.2), we get $A(-k_m)B(k_m) = 2\,|\,k_m\,|$, which shows that $A(-k_m)$ is real, since $B(k_m) = \overline{B(k_m)}$ by (6.1). Putting

$$(6.4) \qquad S(k) = A(-k)/A(k), \qquad (\Im k \geqq 0, k \neq 0),$$

we obtain therefore

$$\rho_m = (2/\pi)\,|\,k_m/A(-k_m)\,|^2 \oint_{k_m} S(k)\,dk.$$

Again, from (6. 3) follows

$$s_1(x, \lambda_m) = \tfrac{1}{2}A(-k_m)/|k_m|\, u(x,$$

Hence the normalized eigenfunction $u_m(x)$ is represente

$$u_m(x) = -c_m\ (2\pi)^{-\frac{1}{2}}u(x, k_m),\ \text{wher} \qquad k.$$

$S(k)$ will be called the *Heisenberg's S-function,* is a diagonal element of the Heisenberg S-matrix.[25] By virtue of (5. 13), we have $S(k) = e^{2i\delta(k)}$, $(k > 0)$. Hence $S(k)$ is obtained, under the Assumption I, from $e^{2i\delta(k)}$ by analytic c ntinuation, if one knows the "phase shift" $\delta(k)$ from the asymptotic beha of the "normalized eigenfunction" $u_k(x)$. Under he Assumptions $S(k)$ is meromorphic in $\Im k > 0$, has poles $_m = i\,|\,k_m\,|$ $(m =$ of the order 1, and except for these poles, $S(k)$ gular in e $A(k)$ and $A(-k$ are regular in $\Im k > 0$ and om as one sees from (6). Thus the following the strictly founded:

THEOREM 6 BERG).[26] *Let $u_k(x)$, $k > 0$, be the solution of the Schrödinger*

$$d^2u/dx^2 \qquad 1)/x^2 - V(x)]u = 0, \qquad < \infty)$$

satisfying the "boundary-condition"

$$u_k(x) \sim \text{const.}\, x^{\nu+1}, \qquad (x \to 0)$$

and having the asymptotic form

$$u_k(x) \sim (2/\pi)^{\frac{1}{2}} \sin[kx - \tfrac{1}{2}\alpha/k \log x + \delta(k)], \qquad (x \to \infty).$$

Then, under the Assumptions I and II, the function $S(k) = e^{2i\delta(k)}$ can be extended analytically over the whole upper half k-plane. The extended $S(k)$ has, on the imaginary axis, the poles $k_m = i\,|\,k_m\,|$ $(m = 1, 2, \cdots)$ of the order 1, and, except for these poles, $S(k)$ is regular in $\Im k > 0$. The negative eigenvalues of the Schrödinger equation are given by $\lambda_m = k_m{}^2 = -|\,k_m\,|^2$, $(m = 1, 2, \cdots)$ and the corresponding normalized eigenfunctions $u_m(x)$ have the asymptotic expressions:

---

[25] Heisenberg [7]; see also Pauli [10], Ma [9], Bargman [1].

[26] Heisenberg [7], Part III and IV. Cf. also Ma [9], Pauli [10].

$$p\,[-|k_m|\,x-\alpha/2\,|k_m|\log x], \qquad (x\to\infty),$$

$_m$ *are given by* $c_m{}^2 = \oint_{k_m} S(k)\,dk$.

it is an o stion whether the Assumption I is valid for every Schrödinger If Assumption I were not valid, it would be impossibl for $\Im k > 0$. *In case the Assumption* II *is not fulfilled* (wh umption I is valid), $S(k)$ *has, in general, singular points in* $\Im k > 0$ *other than* k_m $(m = 1, 2, \cdots)$, *arising from the numerator* $A(-k)$ *of* $S(k)$. *Thus, in this case, it might be impossible to determine the negative eigenvalues of the Schrödinger equati from the singular points of* $S(k)$.[27] The necessary and sufficient conditio $V(x)$ in order that the Assumptions I and II are valid is not known. an be proved that *the assumptions I and II are valid if the po is analytic in neighborhood of* ∞. This asserts the vali eisenberg's eorem for Schrödinger equat appearing usually i

BIBLIOGRAPHY.

[1] Bargm V., "Remarks on the determination of l field of force from the elastic scattering phase shifts," *Physical Review*, vol. 75 (1949), pp. 301-303.

[2] Bôcher, M., "On regular singular points of linear differential equations of the second order whose coefficients are not necessarily analytic," *Transactions of the American Mathematical Society*, vol. 1 (1900), pp. 40-52.

[3] Hartman, P., "On the spectra of slightly disturbed linear oscillators," *American Journal of Mathematics*, vol. 71 (1949), pp. 71-79.

[4] ——— and Wintner, A., "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[5] ——— and Wintner, A., "On the location of spectra of wave equations," *American Journal of Mathematics*, vol. 71 (1949), pp. 214-217.

[6] ——— and Wintner, A., "A criterion for the non-degeneracy of the wave equation," *American Journal of Mathematics*, vol. 71 (1949), pp. 206-213.

[7] Heisenberg, W., "Die beobachtbaren Grössen in der Theorie der Elementarteilchen," I, *Zeitschrift für Physik*, vol. 120 (1943), p. 513; II, *ibid.*, p. 673. III, IV, unpublished.

[8] Jošt, R., *Helvetica Physica Acta*, vol. 22 (1947), p. 256.

---

[27] S. T. Ma has pointed out this fact by an example. Cf. Ma [9].

[9] Ma, S. T., " On a general condition of Heisenberg *Review*, vol. 71 (1947), pp. 195-200.

[10] Pauli, W., *Meson theory of nuclear forces*, New York (1

[11] Stone, M. H., *Linear transformations in Hilbert space*, Society Colloquium Publications, vol. 15 (1932).

[12] Strutt, M. J. O., *Lamésche, Mathieusche und verwandt en in Physik* *Technik*, Ergebnisse der Mathematik un , vol. 1, no. 3 (1932).

[13] Titchmarsh, E. C., *Eigenfunction expansions associ d-order differential equations*, Oxford (1946).

[14] Weyl, H., " Über gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. (1910), pp. 220-269.

[15] ———, " Über gewöhn e Differentialgleichungen mit singulären Stellen und ihre Eigenfunk n," *Göttinger Nachrichten* (1910), pp. 442-467.

6] ———, " Über anlinnasche Interpolationsproblem und sein infinitesimales *nnals of Mathematics*, vol. 36 (1935), pp. 230-254.

17] W r, A., " A tegrations of the adiab ic oscillator," *American Journal* s, vol. 69 (1947), pp. 272.

[18] ntinuous spectra," *American Journal of Mathe-* ol pp. 22-30.

[19] ———, " On zation of characteristic differentials in continuous spectra *view*, vol. 72 (1947), pp. 516-517.

[20] ———, " Stab trum in the wave mechanics of tices," *Physical Review*), pp. 81-82.

[21] ———. " Asy ation of the adiabatic oscillator in yperbolic range," *Du ematical Journal*, vol. 15 (1948), (IV), 55-67.

ODS FOR THE ISOPERIMETRIC PROBLEM OF ZA IN NON-PARAMETRIC FORM.*

By William T. Reid.

1. Int In a previous paper [15][1] the author derived an effective Lindeberg theorem for isoperimetric problems of Bolza type in non-parametric form, and with the aid of this theorem established sufficient conditions for a strong relative minimum. The a r has felt, however, that the sufficiency theorem of [15] should be deri by an extension of the expansion method of proof that he had used ordinary problems of Bolza in non-parametric form ([12], [13] out the aid of auxiliary "Lindeberg theorem." In Part I resent pa such proof is presented.

Section **2** is concerned with the for while certain preliminary results for the expansion proof are given in Section **3**. In Section **4** the increment of the to be minimized is expanded in terms of arbitrary "slope-functi multipliers" satisfying c itinuity properties. Section **5** the proof of the sufficien itions for a strong relative minim ed in Theorem 2.1. Section **6** tains a brief discussion of an "Osgood theorem" for the problem under consideration.

Theorem 2.1 is equivalent to the sufficiency theorems proved by Hestenes [5] and Reid [15]. It is to be pointed out, however, that for a problem equivalent to the one herein treated, Hestenes ([7], [8]; see, in particular, the statement on page 510 of [8]) has shown that conditions weaker than those of Theorem 2.1 suffice to insure a strong relative minimum. In contrast to the method of the present paper, that of Hestenes is indirect, and uses results which are in the nature of a Lindeberg condition in terms of the Weierstrass $\mathcal{E}$-function as derived by the author [15]. It is to be remarked that for the ordinary problem of Bolza the method of the present paper provides simplification of details in the expansion proofs previously given by the author in [12] and [13].

---

* Received February 26, 1949; presented to the American Mathematical Society, September 4, 1947.

[1] Numbers in square brackets refer to the bibliography at the end of this paper.

The proof of Theorem 3. 3, which is funda proof of Section **5**, is the goal of Part II of the pr has been written in such a manner, however, that to a read independently of Part I. In addition, other results which are of importance in themselves. For example, Th ex earlier results of Radon ([9], [10]) and the the relation between the positiveness of the second variatio em and the existence of certain types of solutions of the as dre matrix differential equation of Riccati type. The indicated Theorem 8. 1 shows that with the aid of a preliminary alteration it is possible to eliminate the condition of non-ta cy used by Hestenes [3] and Bliss [2] in the proof of a corresponding ult. Theorems 11. 1 and 12. 1 are useful in the treatment of certain typ self-adjoint boundary problems.

art I.—Sufficient for a Strong Relative Minimum.

il blem. The problem be considered is that of m ve al

$$(2.1) \qquad J \quad), x_2, y(x_2)) + \int_{x_1}^{x_2} f(x, y, y')\,dx$$

in a class of arc

$$(2.2) \qquad\qquad (i = 1, \cdots, n;\ x_1 \quad \leq x_2),$$

satisfying auxiliary first order differential equations

$$(2.3) \qquad \phi_\alpha(x, y, y') = 0, \qquad (\alpha = 1, \cdots, m < n),$$

the end-conditions

$$(2.4) \qquad \psi_\nu(x_1, y(x_1), x_2, y(x_2)) = 0, \qquad (\nu = 1, \cdots, p \leq 2n + 2),$$

and the isoperimetric conditions

$$(2.5) \qquad J_s \equiv g_s(x_1, y(x_1), x_2, y(x_2)) + \int_{x_1}^{x_2} f_s(x, y, y')\,dx = 0, \qquad (s = 1, \cdots, q).$$

For brevity, this minimum problem will be referred to as "problem B."

It is supposed that there is given an open region $\mathfrak{R}$ of points $(x, y, r) = (x, y_1, \cdots, y_n, r_1, \cdots, r_n)$ in which the functions[2] $f(x, y, r)$, $\phi_\alpha(x, y, r)$,

---

[2] For the analysis of this paper it actually suffices to assume that the functions f, ϕ_α, f_s are of class $C^{(2)}$ in $\mathfrak{R}$. Under this weaker assumption, however, the accessory

$^{3)}$, and the $m \times n$ matrix of partial derivatives
k m. We suppose also that there is an open region
mensional $(x_1, y_{i1}, x_2, y_{i2})$-space in which the functions
$C^{(2)}$, and, moreover, the matrix of partial derivatives

$$\| \psi_{\nu,1} \quad \psi_{\nu,j1} \quad \psi_{\nu,2} \quad \psi_{\nu,j2} \|$$

has rank p. $(x_1, y_{i1}, x_2, y_{i2})$, will be said to be "admissible" if it lies in ectively. Corresponding to the terminology of Bliss [2; p. 194], a 2) whose defining functions are continuous and have piecewise continuous derivatives will be termed admissible if its elements $(x, y(x), y'(x))$, $x_1 \leq x \leq x_2$, and $(x_1, y_i(x_1), x_2$ $_2))$ are admissible.

If we set

$$F(x, y, r, \lambda, \bar{\lambda}) = \lambda_0 f(x, y, r) + \lambda_\alpha \phi_\alpha(x \qquad f_s(x, y, r),$$

where repetition of a subscript in a term deno tion with respect this index on its spe ed range, an extremal efined as a issible arc of class $C^{(2)}$, and set of multipliers λ_0 $C^{(1)}$, $\bar{\lambda}_s =$ constant, such that along E,

$$dF_{r_i}/dx - F_{y_i} = 0,$$

An extre $y_i(x), \lambda_0, \lambda_\alpha(x), \bar{\lambda}_s$ is said to s ltiplier rule with constan the relation

$$[(F - y'_i F_{r_i})\,dx + F_{r_i} dy_i]_1^2 + \lambda_0 dg + \bar{\lambda}_s dg_s + e_\nu d\psi_\nu = 0$$

holds for every choice of the differentials dx_1, dy_{i1}, dx_2, dy_{i2}. An extremal is termed non-singular if the matrix

$$\begin{Vmatrix} F_{r_i r_j} & \phi_{\beta r_i} \\ \phi_{\alpha r_j} & 0_{\alpha\beta} \end{Vmatrix}, \qquad (i, j = 1, \cdots, n;\ \alpha, \beta = 1, \cdots, m),$$

is non-singular along this extremal.

Throughout this paper we shall be concerned with an extremal having the leading multiplier λ_0 *different from zero, and, hence, without loss of generality, taken equal to unity.* Such an extremal E is said to satisfy the Weierstrass condition II_N if N is a $(2n + m + q + 1)$-dimensional neighborhood of the elements $(x, y_i(x), y'_i(x), \lambda_\alpha(x), \bar{\lambda}_s)$ of E such that

---

differential equations may not be expressed as linear equations of the second order, but may be written in canonical form.

$$(2.6)\quad \mathcal{E}(x, y, r, \lambda, \bar{\lambda}; \tilde{r}) \equiv F(x, y, \tilde{r}, \lambda, \bar{\lambda}) - F(x,$$
$$- (\tilde{r}_i - r_i) F_{r_i}(x, y, r, \lambda,$$

for arbitrary $(x, y, r, \lambda, \bar{\lambda})$ of N and all $\tilde{r} = (\tilde{r}_i)$ suc
$(x, y, \tilde{r})$ are distinct admissible sets which satisfy $\phi_\alpha(x, y, r$
$(\alpha = 1, \cdots, m)$. As usual, the condition obtained II_N by deleting
the equality sign in (2.6) is referred to as I I_N and non-
singularity imply the strengthened Clebsch condit ach element
$(x, y(x), y'(x), \lambda(x), \bar{\lambda})$ of E the quadratic form $F_{r_i r_j}$ greater than
zero for arbitrary sets $(\pi_i) \neq (0_i)$ satisfying $\phi_{\alpha r_j} \pi_j = 0$, $(\alpha = 1, \cdots, m)$.
Moreover, if E is non- lar and satisfies II_N, then it also satisfies II'_N if
N is properly restricted [4]).

In the terminology Bliss [2; p. 195], a pair of constants ξ_1, ξ_2,
gether with an $\equiv (\eta_i(x))$ of functions which are continuous
d have ewise derivatives on $x_1 \leq x \leq x_2$, will be called an
admis iation. uations of variations of (2.3), (2.4), (2.5)
alon d are, respectively,

$$(2.7)\quad \qquad + \phi_{\alpha y_j} \eta_j = 0,$$

$$(2.8)\quad \Psi_\nu(\xi_1, \eta($$
$$\qquad x_1) + \xi_1 y'(x_1), \xi_2, \eta(x_2) + \xi_2 y' \qquad) = 0,$$

$$(2.9)\quad \mathrm{X}_s(\xi_1, \xi_2, \eta \qquad G_s(\xi_1, \eta(x_1), \xi_2, \eta(x_2))$$
$$+ \int_{x_1}^{x_2} (f_{s r_j} \eta'_j + f_{s y_j} \eta_j)\, dx = 0,$$

where in (2.9),

$$(2.10)\quad G_s(\xi_1, \eta(x_1), \xi_2, \eta(x_2))$$
$$= dg_s(\xi_1, \eta(x_1) + \xi_1 y'(x_1), \xi_2, \eta(x_2) + \xi_2 y'(x_2)) + f_s \xi \Big|_1^2 ;$$

in each of the expressions (2.7), (2.8), (2.9), (2.10) the arguments of the partial derivative functions are those given by the elements $(x, y(x), y'(x))$ of C. As in Bliss [2], corresponding numerical subscripts denote the coefficients of the respective variables $\xi_1, \eta_i(x_1), \xi_2, \eta_i(x_2)$ in Ψ_ν and G_s; specifically,

$$(2.11)\quad \Psi_\nu(\xi_1, \eta(x_1), \xi_2, \eta(x_2)) \equiv \Psi_{\nu,1}\xi_1 + \Psi_{\nu,j1}\eta_j(x_1) + \Psi_{\nu,2}\xi_2 + \Psi_{\nu,j2}\eta_j(x_2),$$

with a similar notation for the coefficients of G_s, $(s = 1, \cdots, q)$.

For an extremal E: $y_i(x)$, $\lambda_0 = 1, \lambda_\alpha(x), \bar{\lambda}_s$, $(x_1 \leq x \leq x_2)$, which satisfies the multiplier rule with constants e_ν, the second variation is written

$$(\xi_1, \eta(x_1), \xi_2, \eta(x_2)) + \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')\,dx.$$

$$_{r_i r_j}\pi_i\pi_j + 2F_{r_i y_j}\pi_i\eta_j + F_{y_i y_j}\eta_i\eta_j,$$

$$2\gamma(\quad \eta \quad x_1) \qquad)) = 2H(\xi_1, \eta(x_1) + \xi_1 y'(x_1), \xi_2, \eta(x_2) + \xi_2 y'(x)),$$

where, as a $\quad y_{i1}, dx_2, dy_{i2}),$

$$2H(dx_1, \qquad y_{i2}) = [(F_{\dot x} - y'_i F_{y_i})\,dx^2 + 2F_{y_i} dy_i dx]_1^2 + 2\mathfrak{G},$$

and $2\mathfrak{G}$ is the quadratic form in $(dx_1, dy_{i1}, dx_2, dy_{i2})$ whose coefficients are the respective second order derivatives of $\mathfrak{g} = g \quad {}_{\nu} + \bar\lambda_s g_s$, while in all terms the partial derivatives of F and of $\mathfrak{g}$ are uated along E.

An extremal satisfying the multiplier rule be said to satisfy condition IV also if along this extremal $J_2(\xi, \eta$ arbitr non-nu admissible variations $(\xi, \eta) = (\xi_1, \xi_2, \eta_i(x))$ wh y th ondi equations of variatio (2. 7), (2. 8) and (2. s usual, dition obtained from the statement of IV upon by "$J_2(\xi, \eta) > 0$" will be referred to as condi is own, non-singularity and IV imply condition III′.

If we set

$$(2.\,13) \qquad \pi, \mu, \bar\mu) = \omega(x, \eta, \pi) + \mu_a \Phi_a(x, \qquad {}_{sr_j}\pi_j + f_{sy_j}\eta_j),$$

an accesso extremal is defined as a set of functions $\eta_i(x)$ of class $C^{(2)}$, and a set of multipliers $\mu_a(x)$ of class $C^{(1)}$, $\bar\mu_s$ = constant, such that

$$(2.\,14) \qquad (d/dx)\Omega_{\pi_i}(x, \eta, \eta' \mu, \bar\mu) - \Omega_{\eta_i}(x, \eta, \eta', \mu, \bar\mu) = 0, \quad \Phi_a(x, \eta, \eta') = 0.$$

Along a non-singular extremal equations (2. 14) may be written in terms of the canonical variables $(x, \eta_i, \zeta_i = \Omega_{\pi_i}(x, \eta, \eta', \mu, \bar\mu))$ as

$$(2.\,15) \qquad \eta'_i = \mathfrak{H}_{\zeta_i}, \qquad \zeta'_i = -\mathfrak{H}_{\eta_i}$$

where $\mathfrak{H}$ is the corresponding Hamiltonian function involving the parameters $\bar\mu_s$, $(s = 1, \cdots, q)$.

The sufficiency theorem to be proved in this paper is as follows.

THEOREM 2. 1. *Suppose that* E: $y_i(x), \lambda_0 = 1, \lambda_a(x), \bar\lambda_s$, $(x_1 \leq x \leq x_2)$, *is a non-singular extremal satisfying* (2. 4) *and* (2. 5), *and that* E *satisfies with constants* e_ν *the multiplier rule and conditions* II'_N *and* IV'. *Then there exists a neighborhood* $\mathfrak{F}$ *of* E *in xy-space and a neighborhood* $\mathfrak{D}$ *of the*

end-points of E in $(x_1, y_{i1}, x_2, y_{i2})$-space such that
admissible arc C satisfying (2.3), (2.4), (2.5) *wh*
points in $\mathfrak{D}$, and is not identical with E.

3. Preliminary results. As shown by Reid [15; E is an
extremal for the problem B satisfying III′ then out loss
of generality that E satisfies the following long E *the*
quadratic form $F_{r_i r_j \pi_i \pi_j}$ is positive definite. More

$$f^*(x, y, r) = f(x, y, r) + \tfrac{1}{2} l \phi_a(x, y, r) \phi_a(x, y, r),$$

and B* denotes the m problem in which (2.3), (2.4), (2.5) are as
in problem B, but (2. altered by substituting f^* for f, then for a
suitable choice of th t l the corresponding quadratic form $F^*_{r_i r_j \pi_i \pi_j}$
r B* is n arc (2.2) clearly affords a minimum for B
and nimum for B*; also, E: $y_i(x), \lambda_0, \lambda_a(x), \bar{\lambda}_s$ is an
extre constants e_ν the mult ier rule for B if and
only, satisfying the corresponding multiplier rule
with the same consta In addition, for an extremal with $\lambda_0 = 1$ each of
the following con ds for problem B if and only if the condition
holds for problem , II′$_N$, non-singularity, III, III′, IV, IV′.

The followin on the behavior of the Weier $\mathcal{E}$-function
involves the non- onotone non-decreasing convex

$$\mathfrak{R}(t) = t^2/(t + 1), \qquad t \geq 0. \tag{3.1}$$

As noted in Reid [15; pp. 681, 683], this function has the following properties:

$$(3.2)\quad \begin{array}{lll} \text{(i)} & \mathfrak{R}(t) \leq \operatorname{Min}(t, t^2) \leq 2\mathfrak{R}(t), & (t \geq 0); \\ \text{(ii)} & a \operatorname{Min}(1, a)\mathfrak{R}(t) \leq \mathfrak{R}(at) \leq a \operatorname{Max}(1, a)\mathfrak{R}(t), & (a \geq 0,\ t \geq 0); \\ \text{(iii)} & \mathfrak{R}(t_1 + t_2) \leq 2(\mathfrak{R}(t_1) + \mathfrak{R}(t_2)), & (t_1 \geq 0, t_2 \geq 0). \end{array}$$

From the Corollary to Theorem 3.2 and Theorem 3.3 of Reid [15] one obtains the following result.[3]

THEOREM 3.1. *If E: $y_i(x), \lambda_0 = 1, \lambda_a(x), \bar{\lambda}_s$ is an extremal for a*

---

[3] Since for admissible sets (x, y, r), $(x, y, \bar{r})$ satisfying $\phi_a(x, y, r) = 0 = \phi_a(x, y, \bar{r})$ the Weierstrass function for B is equal to the Weierstrass function for B^*, while non-singularity and II_N imply III' as pointed out above, we have that inequalities (3.3), (3.4) contain a proof of the previously stated result that non-singularity and II_N imply II'_N if N is properly restricted.

[illegible]*s conditions* III* *and* II_N, *then there exists a neigh-* [illegible]*ents of E in $(x, y, r, \lambda, \bar{\lambda})$-space, and positive constants* [illegible] *$(x, y, r, \lambda, \bar{\lambda})$ is in $\mathfrak{N}$ then (x, y, r) is admissible, and if* [illegible]*t such that $(x, y, \tilde{r})$ is admissible and $\phi_a(x, y, \tilde{r}) = 0$, then*[4]

$$(3.3) \qquad [illegible] \geq \kappa \mathfrak{N}(\| \tilde{r} - r \|),$$

$$(3.4) \qquad | [illegible] \leq \kappa_0 \mathcal{E}(x, y, r, \lambda, \bar{\lambda}; \tilde{r}), \qquad (s = 1, \cdots, q),$$

where $\mathcal{E}_s(x, y, \tilde{r}; r) \equiv f_s(x, y, \tilde{r}) - f_s(x, y, r) - (\tilde{r}_i - r_i) f_{sr_i}(x, y, r)$, *the ordinary Weierstrass function for* $f_s(x, y, r)$.

Using the above inequalities (3.2), one [illegible] establish the following integral inequality which will be em[illegible] i[illegible] proof of the sufficiency theorem. The proof may be obtained by [illegible] sa[illegible] of argument as given for Theorem 5.1 of Reid [12], (see also Lem[illegible]).[5]

THEOREM 3.2. *If* $\boldsymbol{h}_\tau(x)$, $(\tau = 1, \cdots, n)$ [illegible]*inuous functions on* $X_1 \leq x$ [illegible] X_2, *and* $\| \boldsymbol{h}(x) \| \leq \delta$ [illegible]

$$(3.5) \qquad \int_{X_1}^{X_2} \| \boldsymbol{h}(x) \| \cdot \| \boldsymbol{h}'(x) \| \, dx \leq d_1 \Big[\int_{X_1}^{X_2} \mathfrak{N}(\| [illegible] \, dx + \| \boldsymbol{h}(X_1) \|^2 \Big],$$

$$(3.6) \qquad \int_{X_1}^{X_2} \| \boldsymbol{h}(x) \|^2 dx \leq d_2 \Big[\int_{X_1}^{X_2} \mathfrak{N}(\| \boldsymbol{h}'(x) [illegible] X_1) \|^2 \Big],$$

where d_1 [illegible] $\mathrm{Max}(1, 2\delta)\, \mathrm{Max}(1, X_2 - X_1)$, $d_2 = [illegible] (X_2 - X_1)$.

The following result for the second variation, which is fundamental for the expansion proof of Theorem 2.1, is a direct consequence of Theorem 12.2. For an indication of the relation of this theorem to previous results of Hestenes and Bliss the reader is referred to Section 7 and the remark following the statement of Theorem 8.1.

THEOREM 3.3. *If* E: $y_i(x), \lambda_0 = 1, \lambda_a(x), \bar{\lambda}_s$, $(x_1 \leq x \leq x_2)$, *is a non-singular extremal for problem* B *which satisfies with constants* e_v *the multiplier rule and condition* IV′, *then there are constants* $\delta_0 > 0$, $c \geq 0$, $\kappa_1 > 0$ *such that on* $x_1 - \delta_0 \leq x \leq x_2 + \delta_0$ *there exists a non-singular matrix*

---

[4] If for an arbitrary integer k we have defined a k-tuple of real numbers $u = (u_1, \cdots, u_k)$ the symbol $\| u \|$ is used to denote the non-negative square root of $u^2_1 + \cdots + u^2_k$. In particular, this notation is extended to functionals (2.7), (2.8), (2.9); for example, in (3.9), $\| \Psi \|$ is the non-negative square root of $\Psi^2_1 + \cdots + \Psi^2_p$.

[5] It is to be remarked that the statement of Lemma 4.1 of Reid [15] contains two errors: one in the omission of brackets in formulas (4.1) and (4.2); the other in the stated value of the constant d_2.

$\boldsymbol{U}(x) = \| \boldsymbol{U}_{\sigma\tau}(x) \|$, $(\sigma, \tau = 1, \cdots, 2n+q)$, *with el* *continuous functions* $\boldsymbol{\mu}_{a\tau}(x)$, $(\alpha = 1, \cdots, m; \tau = 1,$ *fying the differential equations*

$$(3.7) \qquad \boldsymbol{U}'_{n+i,\tau} = 0, \ \boldsymbol{U}'_{2n+s,\tau} + f_{sr_j}\boldsymbol{U}'_{j\tau} + f_{sy_j}\boldsymbol{U}_{j\tau} =$$

$$(i \qquad , q),$$

and with the property that if the components of η *re absolutely continuous on* $X_1 \leq x \leq$ $_2$, *where* $| X_1 - x_1 | < \delta_0$, $x_2 | < \delta_0$, *and* $\boldsymbol{h}(x) = (\boldsymbol{h}_\tau(x))$ *is def* *y*

$$\boldsymbol{U}_{i\tau}(x)\boldsymbol{h}_\tau(\qquad \boldsymbol{U}_{n+i,\tau}(x)\boldsymbol{h}_\tau(x) = \eta_i(X_2),$$

$$(3.8) \qquad \boldsymbol{U}_{2n+s,\tau}(x)\boldsymbol{h}_\tau(\qquad + f_{sy_j}\eta_j)\,dx, \quad (X_1 \leq x \leq X_2),$$

en

$$\gamma(\xi_1, \qquad J^*_2(\eta; X_1, X_2)$$

$$(3. \qquad \| \boldsymbol{h}(X_1) \|^2 + \| \boldsymbol{h}(X_2 \|^2)$$

$$(\xi_1, \eta(X_1), \xi_2, \eta(X_2)) \|^2 + \| \mathrm{X}(\xi_1, \xi_2, \eta; X_1, X_2) \|^2),$$

where

$$(3.10) \qquad \boldsymbol{\pi}_i = \qquad _\tau(x)\boldsymbol{h}_\tau(x), \quad \boldsymbol{\mu}_a = \boldsymbol{\mu}_a(x; \eta) \equiv \qquad (x)\boldsymbol{h}_\tau(x),$$

$$(3.11) \qquad \boldsymbol{J}^*_2(\eta; X_1,$$

$$= 2 \int_{X_1}^{X_2} \{ \Omega(x, \eta, \boldsymbol{\pi}, \boldsymbol{\mu}, 0) + (\eta'_i - \boldsymbol{\pi}_i)\Omega_{\pi_i}(x, \eta, \boldsymbol{\pi}, \boldsymbol{\mu}, 0) \}\,dx.$$

4. Expansion of the increment of $J(C)$. Suppose that $E: y_i(x), \lambda_0 = 1, \lambda_a(x), \bar{\lambda}_s, (x_1 \leq x \leq x_2)$, is a non-singular extremal for problem B satisfying (2.4), (2.5), as well as the multiplier rule with constants e_ν. Since E is non-singular, by the existence theorem for differential equations there exists a $\delta' > 0$ such that an extension of E is defined and non-singular on the interval $x_1 - \delta' \leq x \leq x_2 + \delta'$. If $0 < \delta < \delta'$, we shall denote by $\mathfrak{F}_\delta$ the neighborhood of E in (x, y)-space consisting of all sets (x, y) satisfying $\| y - y(x) \| < \delta$, $x_1 - \delta < x < x_2 + \delta$. The portion of $\mathfrak{F}_\delta$ in which either $x_1 - \delta < x < x_1 + \delta$ or $x_2 - \delta < x < x_2 + \delta$ defines a neighborhood of the end-points of E in $(x_1, y_{i1}, x_2, y_{i2})$-space which will be denoted by $\mathfrak{D}_\delta$. For the neighborhoods $\mathfrak{F}_\delta$ and $\mathfrak{D}_\delta$ used in the following, it is understood that they are defined as above, with $0 < \delta < \delta'$.

For the set of continuous arcs C: $Y_i(x)$, $(X_1 \leq x \leq X_2)$, where

$+ \delta'$, the symbol $| C - E |$ will denote the "distance
efined by

$$= | X_1 - x_1 | + | X_2 - x_2 | + \operatorname{Max}_{X_1 \leqq x \leqq X_2} \| Y(x) - y(x) \|.$$

Suppose hat fo h arcs C there are defined two functionals $\mathfrak{f}_1(x; C)$ and $\mathfrak{f}_2(x; C)$, We write $\mathfrak{f}_1(x; C) = o\{\mathfrak{f}_2(x; C)\}$ if for each $\epsilon > 0$ there ach that $| \mathfrak{f}_1(x; C) | \leqq \epsilon | \mathfrak{f}_2(x; C) |$, $X_1 \leqq x \leqq X_2$, for all curves , $X_1 \leqq x \leqq X_2$ with $| C - E | < \delta_\epsilon$. Correspondingly, the condition that there exist positive constan such that $| \mathfrak{f}_1(x; C) | \leqq M | \mathfrak{f}_2(x; C) |$, $X_1 \leqq x \leqq X_2$, for all a $E | < \delta$ is written $\mathfrak{f}_1(x; C) = O\{\mathfrak{f}_2(x; C)\}$. The above particular functional $\mathfrak{f}_2(x; C) \equiv 1$ are indicated by d $\mathfrak{f}_1(x; C) = O\{1\}$, respectively.

For the general expansion of O ow it supposed that for each admissible arc C: Y_i wi $x_1 - \delta' < X_1, X_2 < x_2 \ \delta'$, there exist "slop C)), $(i = 1, \cdots, n)$, and "multipliers" $\Lambda \equiv (\Lambda_a$ m), which as functions of x are continuous on $X_1 \leqq x \leqq$ reduce to $y_i(x)$ and $\lambda_a(x)$, respectively, if $C \equiv E$, and as functiona e continuous with respect to the metric (4. 1) in the sense that

$$) - y'(x) \| = o\{1\}, \qquad \| \Lambda(x; C \quad x) \| = o\{1\}.$$

In the pro of the sufficiency theorem as completed in the following section explicit ues of $\mathfrak{p}_i(x; C)$ and $\Lambda_a(x; C)$ will be specified. For brevity, the notation

$$\Delta g = g(X_1, Y(X_1), X_2, Y(X_2)) - g(x_1, y(x_1), x_2, y(x_2))$$

is introduced, with the similar meanings for Δg_s and $\Delta \mathfrak{g}$ where, as in Section 2, $\mathfrak{g} = g + e_\nu \psi_\nu + \bar{\lambda}_s g_s$, and e_ν, $\bar{\lambda}_s$ belong to the set with which E satisfies the multiplier rule.

Now if C: $Y_i(x)$, $x_1 - \delta' < X_1 \leqq x \leqq X_2 < x_2 + \delta'$, is an admissible arc satisfying (2. 3), (2. 4), (2. 5),

$$\begin{aligned} \Delta J &= \Delta g + \int_{X_1}^{X_2} f(x, Y, Y')\, dx - \int_{x_1}^{x_2} f(x, y, y')\, dx \\ &= \Delta \mathfrak{g} + J^0 - J^1 - J^2, \end{aligned} \tag{4. 2}$$

where

$$J^0 = \int_{X_1}^{X_2} \{ F(x, Y, Y', \Lambda, \bar{\lambda}) - F(x, y, y', \lambda, \bar{\lambda}) \} dx, \tag{4. 3}$$

$$J^k = (-1)^k \int_{X_k}^{x_k} F(x, y, y', \lambda, \bar{\lambda})\, dx, \tag{4.4}$$

Writing $\eta_i(x) \equiv \eta_i(x; C) = Y_i(x) - y_i(x)$, $\mu_a(x) \equiv \mu_a$
$- \lambda_a(x)$, $(X_1 \leq x \leq X_2)$, and using the expansion metho
Reid [12], it follows that

(4.5)

$$J^0 = \int_{X_1}^{X_2} \mathcal{E}(x, Y, \mathfrak{p}, \Lambda, \bar{\lambda}; Y')\, dx + \int_{X_1}^{X}$$

$$+ \int_{X_1}^{X_2} \{\Omega($$ $$, \mu, 0) + (Y'_i - \mathfrak{p}_i)\Omega_{\pi_i}(x, \eta, \mathfrak{p} - y', \mu, 0)\}\, dx$$

$$+ \int_{X_1}$$ $$\mathfrak{p}_i)\mathfrak{b}_i(x; C)\}\, dx,$$

where for a given C) and $\mathfrak{b}_i(x; C)$ are continuous in x
$X_1 \leq x$

(4.6)

$$\{\| \eta(x) \|^2 + \| \mathfrak{p}(x) - y'(x) \|^2 + \| \mu(x) \|^2\},$$

$$\eta(x) \| + \| \mathfrak{p}(x) - y' \quad) \| + \| \mu(x) \|\}.$$

It is to be understo at the arguments of the partial derivatives of F occurring in the se third integrals of (4.5) are the elements of the extremal E.

In view of the agrange equations the second i n (4.5) is equal to $F_{r_j}(X_2)\eta_j(X_2) - F_{r_j}(X_1)\eta_j(X_1)$. Moreover, by pansion process of Section 6 of Reid [12], it follows that

$$\Delta \mathfrak{g} - J^1 - J^2 + F_{r_j}\eta_j \Big|_{X_1}^{X_2} = \gamma(X_1 - x_1, \eta(X_1), X_2 - x_2, \eta(X_2)) + \gamma^*(C), \tag{4.7}$$

where for each $\epsilon > 0$ there is a $\delta_\epsilon > 0$ such that if the end-points of C are in the $\mathfrak{D}_{\delta_\epsilon}$ neighborhood of the ends of E, then $| \gamma^*(C) | \leq \epsilon\{(X_1 - x_1)^2 + (X_2 - x_2)^2 + \| \eta(X_1) \|^2 + \| \eta(X_2) \|^2\}$. In particular,

$$\gamma^*(C) = o\{(X_1 - x_1)^2 + (X_2 - x_2)^2 + \| \eta(X_1) \|^2 + \| \eta(X_2) \|^2\}. \tag{4.8}$$

Combining (4.5) and (4.7), it results that if C: $Y_i(x)$, $x_1 - \delta' < X_1 \leq x \leq X_2 < x_2 + \delta'$, is an admissible arc satisfying (2.3), (2.4), (2.5), then

$$\Delta J = \int_{X_1}^{X_2} \mathcal{E}(x, Y, \mathfrak{p}, \Lambda, \bar{\lambda}; Y')\, dx + \tfrac{1}{2}\hat{J}_2 + \int_{X_1}^{X_2} [\mathfrak{a}(x; C) + (Y'_i - \mathfrak{p}_i)\mathfrak{b}_i(x; C)]\, dx + \gamma^*(C), \tag{4.9}$$

) and $\gamma^*(C)$ are "remainder" expressions possessing (4.6) and (4.8), respectively, and

$$_1 - x_1, \eta(X_1), X_2 - x_2, \eta(X_2))$$

$$+2 \quad (x, \eta, \mathfrak{p} - y', \mu, 0) + (Y'_i - \mathfrak{p}_i)\Omega_{\pi_i}(x, \eta, \mathfrak{p} - y', \mu, 0)]dx.$$

As in th proofs of the author ([12], [13]), it will be proved that able choices of $\mathfrak{p}_i$ and $\Lambda_a(x; C)$ the first two terms of (4.9) are the dominant ones in ion.

5. Proof of Theorem 2 nts at the beginning of Section **3** we may assume es condition III* in addition to the stated hypothe ch assumption will be made in this and the followi tive a not greater than either the δ_0 of Theorem 3. n th first paragraph of S ion **4**, and consider (x), $(X_1 \leq x \leq X_2)$, for problem B which satisfies and has end-points in the $\mathfrak{D}_\delta$ neighborhood of the points of E. Let $\eta_i(x) = Y_i(x) - y_i(x)$, $X_1 \leq x \leq X_2$, and co e functions $\boldsymbol{h}(x) = (\boldsymbol{h}_\tau(x))$, $(\tau = 1, \cdots, 2n + q)$, defined by (3.8) of Theorem 3.3. Fo nissible arc C these functions are ous and have piecewise co s derivatives; indeed, if the functions $Y_i(x)$ defining C are merely ab itely continuous then the functions $\boldsymbol{h}_\tau(x)$ are absolutely continuous on $X_1 \leq x \leq X_2$. Clearly $\| \eta(x) \| = O\{\| \boldsymbol{h}(x)\}\|$; moreover, the $(2n + q)$-dimensional vector

$$\boldsymbol{\eta}(x) \equiv (\boldsymbol{\eta}_\sigma(x)) \equiv \left(\eta_i(x), \eta_i(X_2), \int_x^{X_2} (f_{sr_j}\eta'_j + f_{sy_j}\eta_j)\, dx\right)$$

satisfies the folowing relations: $\| \boldsymbol{h}(x) \| = O\{\| \boldsymbol{\eta}(x) \|\}$, and $\| \boldsymbol{\eta}(x) \| = O\{\| \boldsymbol{h}(x) \|\}$. Also, since upon integration by parts,

$$\boldsymbol{\eta}_{2n+s}(x) = f_{sr_j}\eta_j \Big|_x^{X_2} + \int_x^{X_2} (f_{sy_j} - (d/dt) f_{sr_j})\eta_j\, dt,$$

it follows that

$$\| \boldsymbol{\eta}(x) \| = O\{\mathrm{Max}_{X_1 \leq x \leq X_2} \| \eta(x) \|\}; \tag{5.1}$$

in particular, as functionals of C, $\| \boldsymbol{\eta}(x) \| = o\{1\}$, and $\| \boldsymbol{h}(x) \| = o\{1\}$.

Now define functions $\mathfrak{p}_i(x; C)$ and $\Lambda_a(x; C)$ as

$$\begin{aligned} \mathfrak{p}_i(x; C) &= y'_i(x) + \boldsymbol{U}'_{i\tau}(x)\boldsymbol{h}_\tau(x) = y'_i(x) + \boldsymbol{\pi}_i(x; \eta), \\ \Lambda_a(x; C) &= \lambda_a(x) + \boldsymbol{\mu}_{a\tau}(x)\mathbf{h}_\tau(x) = \lambda_a(x) + \boldsymbol{\mu}_a(x; \eta), \end{aligned} \tag{5.2}$$

in the notation of (3.10). These functions sa[illegible]y th
in Section **4**, since for each admissible arc C with end-p
continuous on $X_1 \leq x \leq X_2$, while as functionals of C,

$$(5.3) \quad \| \mathfrak{p}(x;C) - y'(x) \| = O\{\| \boldsymbol{h}(x) \|\},$$

$$\| \Lambda(x; C \quad) \| = O$$

From the definit[illegible] 2), and the [illegible] n equations of
(3.8), it follows that [illegible] hence

$$(5.4) \quad \boldsymbol{h}'(x) \|\}.$$

Moreover, in view o [illegible] ${}_{n+i,\tau}\boldsymbol{h}'_\tau = 0, \ (i = 1, \cdots, n),$
${}_{2n+s,\tau}\boldsymbol{h}'_\tau =$ [illegible], and consequently

$$(5.5) \quad \| = O\{\| Y'(x) - \mathfrak{p}(x;C) \|\}.$$

[illegible] (2.4), upon expanding the functions of these
conditions to first or[illegible] terms it follows that

$$\| \Psi(X_1 - x_1, \quad - x_2, \eta(X_2)) \|$$
$$= \quad - x_1 | + | X_2 - x_2 | + \| \boldsymbol{h}(X_1) \quad (X_2) \|\}.$$

In view of (5.1), it follows immediately that

$$\| \mathrm{X}(X_1 - x_1, X_2 - x_2, \eta; X_1, X_2) \| = o\{1\}.$$

Moreover, since both C and E satisfy (2.5), upon expanding $\Delta J_s = J_s(C) - J_s(E)$ in the manner analogous to the expansion of ΔJ in Section **4**, and using the above stated order relations for $\| \eta \|$, $\| \mathfrak{p} - y' \|$ and $\| Y' - \mathfrak{p} \|$, it results that

$$\mathrm{X}_s(X_1 - x_1, X_2 - x_2, \eta; X_1, X_2) = - \int_{X_1}^{X_2} \mathcal{E}_s(x, Y, \mathfrak{p}; Y')\, dx + \mathrm{X}^*_s(C),$$

with

$$\mathrm{X}^*_s(C) = O\{(X_1 - x_1)^2 + (X_2 - x_2)^2 + \| \boldsymbol{h}(X_1) \|^2 + \| \boldsymbol{h}(X_2) \|^2$$
$$+ \int_{X_1}^{X_2} (\| \boldsymbol{h}(x) \|^2 + \| \boldsymbol{h}(x) \| \ \| \boldsymbol{h}'(x) \|)\, dx\}.$$

Consequently, by Theorem 3.3 the functional $\hat{J}_2$ of (4.9) satisfies the inequality

$$x_1 \ldots X_2 - x_2)^2 + \| \boldsymbol{h}(X_1) \|^2 + \| \boldsymbol{h}(X_2) \|^2]$$

$$\sum_{s=1} \int_{X_1}^{X_2} | \mathcal{E}_s(x, Y, \mathfrak{p}; Y') | \, dx\}$$

$$\{(X_1 - x_1)^2 + (X_2 - x_2)^2 + \| \boldsymbol{h}(X_1) \|^2 + \| \boldsymbol{h}(X_2) \|^2\}$$

$$X_2 \ldots \|^2 + \| \ldots) \|) \, dx\}.$$

Now for $|C - E|$ sufficie … $(x, Y(x); \mathfrak{p}(x; C), \Lambda(x; C), \bar{\lambda})$ of the integran … n (4.9) lie in the neighborhood $\mathfrak{N}$ of Theorem … rem 3.1, together with (5.5) and (3.2ii), th … δ'' such that if C: $Y_i(x)$, $(X_1 \leq x \leq X_2)$, i … hich $|C - E| < \delta''$, then

$$\int_{X_1}^{X_2} \mathcal{E}(x, Y, \mathfrak{p}, \Lambda, \ldots \leq \kappa \ldots \int_{X_1} \tag{5.7}$$

From the expressi… (4.9) for ΔJ, togeth … 6), (5.7), the order relations derived above, inequalit … (4) … 3.1, and Theorem 3.2, it follows that if $0 < \kappa'' < \kappa'$, … $< \kappa_1/2$, then there is a $\delta > 0$ such that if C: $Y_i(x)$, $(X_1 \leq x \leq$ … an admissible arc satisfying (2.… (2.4), (2.5), and for which … $< \delta$, then

$$\Delta J \ldots \kappa'' \int_{X_1}^{X_2} \mathfrak{N}(\| \boldsymbol{h}'(t) \|) \, dt \tag{5.8}$$
$$+ \kappa'_1[(X_1 - x_1)^2 + (X_2 - x_2)^2 + \| \boldsymbol{h}(X_1) \|^2 + \| \boldsymbol{h}(X_2) \|^2].$$

Consequently, for such an admissible arc C we have $J(C) \geqq J(E)$, and the equality holds if and only if $X_k = x_k$, $\| \boldsymbol{h}(X_k) \| = 0$, $(k = 1, 2)$, and the integral in (5.8) is equal to zero, in which case $\boldsymbol{h}_\sigma(x) \equiv 0$ and $C \equiv E$. In particular, the conclusion of Theorem 2.1 holds for $\mathfrak{F} = \mathfrak{F}_\delta$ and $\mathfrak{D} = \mathfrak{D}_\delta$, with δ suitably small. It is to be remarked that this result remains valid for arcs C: $Y_i(x)$, $(X_1 \leq x \leq X_2)$, lying in $\mathfrak{F}$ and having end-points in $\mathfrak{D}$ which satisfy (2.4), while $Y_i(x)$, $(i = 1, \cdots, n)$, are absolutely continuous functions which satisfy $\phi_\alpha(x, Y, Y') = 0$ almost everywhere on $X_1 \leq x \leq X_2$, are such that the integrals in (2.1) and (2.5) exist in the Lebesgue sense, and which satisfy (2.5).

6. An Osgood theorem. In view of the preceding expansion proof of Theorem 2.1, the following Osgood theorem for problem B may be established directly, and without recourse to an ... Bolza, as was done in Section 6 of Reid [15].

THEOREM 6.1. *Suppose that E is an extremal for*
satisfies the hypotheses of Theorem 2.1. Then there exists
$\mathfrak{F}$ of E in xy-space and ... rhood $\mathfrak{D}$ ... the ... points of E in
$(x_1, y_{i1}, x_2, y_{i2})$-space w ... corresponding
to each neighborhood ... each ... hood $\mathfrak{D}'$ of the
end-points of E inte ... constant $\bar{\kappa} > 0$ such that for
every admissible arc ... , (2.5) which lies in $\mathfrak{F}$ and
has end-points i ... $\mathfrak{F}'$ and have end-points in $\mathfrak{D}'$,
the inequality $J(C)$

Let $\mathfrak{F} = \mathfrak{F}$ ... inequality (5.8) holds for
bitrar ... .4), (2.5), and which lie in
$\mathfrak{F}_\delta$ a ... $\mathfrak{D}_\delta$. For g ... $\mathfrak{F}'$ and $\mathfrak{D}'$ interior to $\mathfrak{F}$ and $\mathfrak{D}$,
resp ... $0 < \bar{\delta} < \delta$, such that ... is interior to $\mathfrak{F}'$ and
$\mathfrak{D}_{\bar{\delta}}$ ... sequently, if C is in $\mathfrak{F}$ with end-points in $\mathfrak{D}$, but
does not lie in $\mathfrak{F}'$ ... have end-points in $\mathfrak{D}'$, then either $(X_1 - x_1)^2$
$+ (X_2 - x_2)^2 \geq \bar{\delta}$ ... re is an x_0 on $X_1 \leq x \leq X_2$ such that $\| \eta(x_0) \|$
$= \| Y(x_0) - y(x_0$ ... In the first case, it follows from (5.8) that
$J(C) - J(E) \geq \kappa'_1$ ... Since $\| \eta(x) \| = O\{\| \boldsymbol{h}(x) \|\}$, the ... constant $M_0 > 0$ such that $\| \eta(x) \| \leq M_0 \| \boldsymbol{h}(x) \|$, $(X_1 \leq x \leq X_2)$, f ... arbitrary admissible arc C: $Y_i(x)$, $(X_1 \leq x \leq X_2)$, lying in $\mathfrak{F}_\delta$; in ... rticular, if $\| \eta(x_0) \| \geq \bar{\delta}$, then $\| \boldsymbol{h}(x_0) \| \geq \bar{\delta}/M_0$. Now, as shown in Section 6 of Reid [15], from (3.2ii), (3.2iii), and Jensen's inequality, it follows that for an arbitrary point x_0 on $X_1 \leq x \leq X_2$ we have

$$\| \boldsymbol{h}(X_1) \|^2 + \int_{X_1}^{X_2} \mathfrak{R}(\| \boldsymbol{h}'(t) \|)\, dt \geq d_0 \mathfrak{R}(\| \boldsymbol{h}(x_0) \|),$$

where d_0 is a positive constant such that Min $(1, 1/(X_2 - X_1)) \geq 2d_0$ when $| X_k - x_k | < \delta$, $(k = 1, 2)$. In particular, if there is an x_0 on $X_1 \leq x \leq X_2$ such that $\| \eta(x_0) \| \geq \bar{\delta}$, from (5.8) it results that

$$J(C) - J(E) \geq d_0 \text{ Min } (\kappa'', \kappa'_1) \mathfrak{R}(\bar{\delta}/M_0).$$

Consequently, if C is an admissible arc satisfying (2.3), (2.4), (2.5) which lies in $\mathfrak{F}$ and has end-points in $\mathfrak{D}$, but which does not lie in $\mathfrak{F}'$ and have end-points in $\mathfrak{D}'$, then $J(C) - J(E) \geq \bar{\kappa}$, where $\bar{\kappa}$ is the smaller of the constants $\kappa'_1 \bar{\delta}^2$ and d_0 Min $(\kappa'', \kappa'_1) \mathfrak{R}(\bar{\delta}/M_0)$.

ory Problem.

remarks. In this part we shall consider a minimum ssentially the accessory problem associated with an ordinary roblem za or an isoperimetric problem of Bolza. In particular, we shall establish in heorem 2 a res lies Theorem 3.3, which was fundamen 1 given in Section **5.** Another establ sult, tha in the consideration of boundary value problems. a problem involving separated end-conditions and the results for the more general problem are obt n to one of the type initially considered, usi ion. Theorem 8.1 below is equivalent to one ir [3]; the method of proof here given parallels tha limina modification that enables one the an non-tangency made by him.

Theorem 8.2 gene lizes the results of S 6]. As pointed out by the author [17], priority is o J. Radon the presentation of the Legendre equation for probler agrange type as a matrix differential equation of Riccati type, and damental theorems on the integra n of such equations. In particula esults of Sections 3 and 4 of 6] are contained in the papers [9] and [10] of Radon.

Throu ut this part matrix notation is used extensively. The transpose of a matrix M is denoted by $\tilde{M}$; in particular, vectors η, π, etc. are treated as one-column matrices.

8. Formulation of the problem not involving isoperimetric conditions.

Let $(\xi, \eta) = (\xi_\rho, \eta_i(x))$, $(\rho = 1, \cdots, k;\ i = 1, \cdots, n;\ x_1 \leq x \leq x_2)$, and consider the quadratic functional

$$(8.1) \qquad J_2(\xi, \eta) \equiv 2\gamma(\xi, \eta(x_1), \eta(x_2)) + \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')\, dx,$$

in which $2\gamma(\xi, \eta_1, \eta_2) \equiv 2\gamma(\xi, (\eta_{i1}), (\eta_{i2}))$ is a quadratic form in the $2n + k$ variables ξ_ρ, η_{i1}, η_{i2}, and $2\omega(x, \eta, \pi)$ is a quadratic form in the $2n$ variables η_i, π_i with coefficients which are functions of x. The symbol B_2 will be used to designate the variational problem involving (8.1) subject to auxiliary first order linear homogeneous differential equations

$$(8.2) \qquad \Phi_\alpha(x, \eta, \eta') \equiv \phi_{\alpha j}(x)\eta'_j + \theta_{\alpha j}(x)\eta_j = 0, \qquad (\alpha = 1, \cdots, m < n),$$

and a set of linear homog us end-co

$$(8.3)\quad \Psi_\nu(\xi, \eta(x_1), \eta(x_2)) \equiv \Psi_{\nu,\rho}\xi_\rho + \Psi_{\nu,j1}\eta_j(x_1) + \Psi_{\nu,j2}$$

$$(\nu = 1, \cdot \cdot$$

It is to be understood th nts of e forms γ and Ψ_ν are real constants; correspo rivatives of γ with respect to are γ_ρ, γ_{i1}, γ_{i2}, respectively. We wri

$$2 \quad)\eta + \tilde{\eta}P(x)\eta,$$

where $R(x)$, $Q(x)$, h $R(x)$ and $P(x)$ symmetric. It is supposed that es $R(x)$, $Q(x)$, $P(x)$, $\phi(x)$ $\| \phi_{\alpha j}($ continuous functions of x nter erval $x_1 \leq x \leq x_2$, while $\phi(x)$ is o rval.

, $(x_1 \leq x \leq x_2)$, will e termed admissible if the ions $\eta_i(x)$, $= 1, \cdot \cdot \cdot, n)$, are continuous and have piecewise continuous derivati $x_1 \leq x \leq x_2$. If $\eta(x)$ is admissible and $\xi = (\xi_\rho)$, $(\rho = 1, \cdot \cdot \cdot, k)$, a rary constants, the set $(\xi, \eta(x))$ will be called admissib; such a l be termed null if $\xi_\rho = 0$, $(\rho = \cdot \cdot, k)$, and $\eta_i(x) \equiv 0$, $(x_1 \leq x \leq x_2;\ i = 1, \cdot \cdot \cdot, n)$. The accessory for an ordinary problem of Bolza is a special case of B$_2$ in which $\xi = \xi_1, \xi_2)$.

If $\Omega(x, \eta, \pi, \mu) = \omega(x, \eta, \pi) + \mu_\alpha \Phi_\alpha(x, \eta, \pi)$, the Euler-Lagrange equations of B$_2$ are

$$(8.4)\quad \begin{aligned} &(d/dx)\Omega_{\pi_i}(x, \eta(x), \eta'(x), \mu(x)) - \Omega_{\eta_i}(x, \eta(x), \eta'(x), \mu(x)) = 0, \\ &\Phi_\alpha(x, \eta(x), \eta'(x)) = 0, \qquad (i = 1, \cdot \cdot \cdot, n;\ \alpha = 1, \cdot \cdot \cdot, m). \end{aligned}$$

If B$_2$ is non-singular, that is, if the $(n + m)$-order square matrix

$$(8.5)\quad \begin{Vmatrix} R(x) & \tilde{\phi}(x) \\ \phi(x) & 0 \end{Vmatrix}$$

is non-singular on $x_1 \leq x \leq x_2$, then in terms of the canonical variables $\eta_i(x)$, $\zeta_i(x) = \Omega_{\pi_i}(x, \eta(x), \eta'(x), \mu(x))$ the system (8.4) may be written

$$(8.6)\quad \eta' = A(x)\eta + B(x)\zeta, \qquad \zeta' = C(x)\eta - \tilde{A}(x)\zeta.$$

For the explicit forms of $A(x)$, $B(x)$, $C(x)$, see, for example, p. 245 of [16]; in particular, the elements of these matrices are continuous and $B(x)$ and

I ξ^* are olutions of (8.6), then
t; w is stant is . hese solutions are said
each other. As shown in Radon [10], or Reid [16], for
roblem B_2 there exists a set of n solutions $\eta_i = U_{ij}(x)$,
$\zeta_i = V_i$, $(j = 1, \cdots, n)$, of (8.6) wi $\equiv \| U_{ij}(x) \|$ non-singular
on $x_1 \leq x \leq x_2$ if only al a continuous solution
$W(x)$ of th a tion"

$$(8.7) \quad W' + WA(x) + \tilde{A}(x) \qquad\qquad , \quad x_1 \leq x \leq x_2.$$

Indeed, if $\eta_i = U_{ij}(x)$, $\zeta_i =$ 6) with $U(x)$
non-singular on $x_1 \leq x \leq x_2$, is a continuous
solution of (8.7); moreover, mutually conjugate
if and only if $W(x) = V(x)$

For B_2 condition III is
be non-negative for arbitrary $\pi =$ satisfying ϕ
III′ is the condition at $\bar{\pi}R(x)\pi > 0$ for ing
$\phi(x)\pi = 0$, $(x_1 \leq x \leq x_2)$. From Radon [10], or hat
if B_2 satisfies condition III′ then $J_2(\xi, \eta) > 0$ for arb y non-null admissible
sets $(\xi, \eta(x))$ satisfying $\xi_\rho = 0$, $\eta_i(x_1) = 0 = \eta_i($ nd only if there is
a continuous symmetric solution of (8.7) on $x_1 \leq$

Corres g to the terminology used in more general variational problems, s said to involve separated end-conditions if upon suitable renumberin f the ξ's there is a k', $(0 \leq k' \leq k)$, such that if we set $\xi^1 = (\xi_{\rho'})$, $(\rho' = 1, \cdots, k')$, and $\xi^2 = (\xi_{\rho''})$, $(\rho'' = k' + 1, \cdots, k)$, then 2γ is of the form $2\gamma^1(\xi^1, \eta(x_1)) + 2\gamma^2(\xi^2, \eta(x_2))$, while the end-conditions (8.3) may be written after possible re-ordering, as two systems of the form

$$(8.3') \qquad \Psi_{\nu'}(\xi^1, \eta(x_1)) = 0, \qquad \Psi_{\nu''}(\xi^2, \eta(x_2)) = 0,$$
$$(\nu' = 1, \cdots, p'; \nu'' = p' + 1, \cdots, p).$$

It is to be remarked that either of the sets ξ^1 or ξ^2 may be non-existent, corresponding to $k' = 0$ and $k' = k$, respectively.

For the general problem B_2 we introduce the notation

$$(8.8) \quad 2\gamma(\xi, \eta(x_1), \eta(x_2); c) = 2\gamma(\xi, \eta(x_1), \eta(x_2)) + c \, \| \Psi(\xi, \eta(x_1), \eta(x_2)) \|^2,$$

with corresponding definitions of $2\gamma^1(\xi^1, \eta(x_1); c)$ and $2\gamma^2(\xi^2, \eta(x_2); c)$ for a problem with separated end-conditions. In particular, we set

$$J_2(\xi, \eta; c) = J_2(\xi, \eta) + c \, \| \Psi(\xi, \eta(x_1), \eta(x_2)) \|^2.$$

ing theorem is

involves separated end
ndition for $J_2(\xi,\eta)$ to be
ing 2), (8.3)
), $\xi_i = V_{ij}(x)$
on $x_1 \leq x \leq x_2$, and which
qualities

$$) - \tilde{a}\bar{U}(x_1)V(x_1)a > 0,$$

$$) + \tilde{b}\bar{U}(x_2)V(x_2)b > 0,$$

$$(\xi_{\rho'}, a_i), \quad (\xi^2, b) = (\xi_{\rho''}, b_i).$$

eor s of quadratic forms (see, for
ex equalities (8. and (8 10) are equivalent to
cor qualities used by Hes nes [3] and Bliss [2;
p. 2 poi Section **7**, however, the proofs of these corresponding
inequalities by Hest nd Bliss involve an additional hypothesis of non-
tangency. From th rk following equation (8.7) it follows that under
the substitution $a^1 =$ $x_1)a$, $a^2 = U(x_2)b$ the result of th above theorem
is expressed in terms of solutions of the Legendre matrix dif equation
(8.7) as follows.

THEOREM 8.2. *Under the hypotheses of Theorem 8.1, a necessary and sufficient condition for $J_2(\xi,\eta)$ to be positive for arbitrary non-null admissible $(\xi,\eta(x))$ satisfying (8.2), (8.3) is that there exist on $x_1 \leqq x \leqq x_2$ a continuous symmetric solution $W(x)$ of (8.7) which satisfies with a suitable constant $c \geqq 0$ the inequalities*

$$(8.11) \qquad 2\gamma^1(\xi^1, a^1; c) - \tilde{a}^1 W(x_1) a^1 > 0,$$

$$(8.12) \qquad 2\gamma^2(\xi^2, a^2; c) + \tilde{a}^2 W(x_2) a^2 > 0,$$

for respective arbitrary non-null sets (ξ^1, a^1), (ξ^2, a^2).

9. An auxiliary lemma. Preliminary to the proof of Theorem 8.1 we shall establish the following result, which holds for the general problem B_2 without the restriction of separated end-conditions.

LEMMA 9.1. *If B_2 is non-singular, and $J_2(\xi,\eta) > 0$ for arbitrary non-*

$(\xi, \eta$

enables one to

the end-conditions (8.3).
that ther s. no
$\eta(x) \not\equiv 0$ on x_1
$x_1 \leq x \leq x_2$ is r_0, and η_i
linearly independent solution
$s = 1, \cdots, r_0$), then the $2n$

$$(9.1) \qquad \left\| \begin{matrix} \eta_{it}(x_1) \\ \eta_{it}(x_2) \end{matrix} \right\|,$$

has rank $2n - r_0$. In partic
all non-null sets of constants $(\xi_\rho,$ uch that

$$(9.2) \qquad \xi_\rho, \eta_i(x) = \eta_{it}(x) z_t, \qquad (\rho$$

satisfies $\| \Psi(\xi, \eta(x_1), \eta(x_2)) \|^2 = 0$ we have $J(\xi$ > 0. Hence by the theorem on quadratic forms of Reid [14] there is nt $c \geq 0$ such that $J_2(\xi, \eta; c) > 0$ for all $(\xi, \eta(x))$ of the form (9.2) ith non-null (ξ_ρ, z_t). Finally, he hypotheses of the lemma, if $(\xi, \eta(x))$ is admissible and (8.2) hol is well-known (see, for example, Bliss [2], p. 233) that there is a unique t of constants z_t such that $\eta^*_i(x) = \eta_{it}(x) z_t$ satisfies $\eta^*_i(x_1) = \eta_i(x_1)$, $\eta^*_i(x_2) = \eta_i(x_2)$. Since $\eta^*_i(x)$, $\zeta^*_i(x) = \zeta_{it}(x) z_t$ is a solution of (8.6) it then follows that

$$J_2(\xi, \eta; c) = J_2(\xi, \eta^*; c) + J_2(0, \eta - \eta^*; c) \geq J_2(\xi, \eta^*; c),$$

and the equality holds only if $\eta(x) \equiv \eta^*(x)$ on $x_1 \leq x \leq x_2$. However, $(\xi, \eta^*(x))$ is of the form (9.2) and $\eta^*(x)$ satisfies (8.2), so that $J_2(\xi, \eta^*; c) \geq 0$, and the equality holds only if $(\xi, \eta^*(x))$ is null. Consequently, the result of the lemma holds for c determined as above.

10. Proof of Theorem 8.1. Let c be a constant such that the result of Lemma 9.1 holds for

$$(10.1) \qquad J_2(\xi, \eta; c) \equiv 2\gamma^1(\xi^1, \eta(x_1); c) + 2\gamma^2(\xi^2, \eta(x_2); c) + \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')\, dx.$$

parti

ystem o

$(\rho' =$ $, p')$,

either the set ξ^2 is $= (\xi_{\rho''})$, $(\rho'' = p' + 1,$ (x_2) the system

$$(\rho'' = p' + 1, \cdots, p),$$

e following we shall suppose lifications which occur when

(86 Bliss [2], let the columns of U $= \| V^1_{ij}(x) \|$ be soluti s η_i, ζ_i of (8.6) such tha form with corresponding constants $\xi_{\rho'} = \xi_{\rho' j}$ a set of n linearly indepen solutions of the algebraic equations

$$(10.4) \qquad \gamma^1_{\rho'}(\xi^1, \eta(\quad) = 0, \qquad -\zeta_i(x_1) + \gamma^1_{i1}(\xi^1, \eta(x_1); c) = 0.$$

It follows immediately that $U^1(x_1)$ is non-singular, and th $U^1_{ij}(x)$, $\zeta_i = V^1_{ij}(x)$, $(j = 1, \cdots, n)$, are mutually conjugate soluti of (8.6). Also $U^1(x)$ is non-singular on $x_1 \leq x \leq x_2$, since otherwise there would exist a solution $\eta^*(x)$, $\zeta^*(x)$ of (8.6) with $\eta^*(x_1) \neq 0$, which satis s (10.4) with suitable ξ^{1*}, and is such that $\eta^*(x_3) = 0$, where $x_1 < x_3 \leq x_2$. Then $\eta(x) = \eta^*(x)$, $(x_1 \leq x \leq x_3)$, $\eta(x) \equiv 0$ on $x_3 \leq x \leq x_2$, $\xi^1 = \xi^{1*}$, $\xi^2 = 0$, would be a non-null admissible set $(\xi, \eta(x))$ satisfying (8.2), and for which $J_2(\xi, \eta; c) = 0$, contrary to Lemma 9.1. Hence $U^1(x)$ is non-singular on $x_1 \leq x \leq x_2$. Similarly, if the columns of $U^2(x)$, $V^2(x)$ are solutions of (8.6) whose end-values at x_2 form with corresponding $\xi_{\rho''} = \xi_{\rho'' j}$ a set of n linearly independent solutions of the algebraic system

$$(10.5) \qquad \gamma^2_{\rho''}(\xi^2, \eta(x_2); c) = 0, \qquad \zeta_i(x_2) + \gamma^2_{i2}(\xi^2, \eta(x_2); c) = 0,$$

then these solutions of (8.6) are mutually conjugate, and $U^2(x)$ is non-singular on $x_1 \leq x \leq x_2$. For brevity, we introduce the notation Ξ^1 for the $k' \times n$ matrix $\| \xi_{\rho' j} \|$, and Ξ^2 for the $(k - k') \times n$ matrix $\| \xi_{\rho'' j} \|$.

If $x_1 < x_3 < x_2$, and $a \equiv (a_i)$, $b \equiv (b_i)$ are such that

$(x))$ d

$\eta($ $(x)a, \quad x_1 \leq$

is admissible and satisfies (
Theorem 86.1 of Bliss [2],
equal to

(10.7) $\tilde{a}[\tilde{V}^1(x$

and hence (10.7) is positi
Since $U^1(x)$ and $U^2(x)$ are no
$-\tilde{U}^1(x)V^2(x)$, which is necessa onstant ma
this interval. In pa cular, for a suitable his
matrix reduces to the identity matrix, and we shal oice
is made; of course, such a choice of $U^2(x)$, $V^2(x$ tails a corresponding
choice of Ξ^2. With this choice we now define $) = U^1(x) + U^2(x)$, $V(x) = V^1(x) + V^2(x)$. As in Bliss [2; p. 248 t follows readily that the colu (x), $V(x)$ are mutually conjugate solutions of (8.6), and that $U($ non-singular on $x_1 \leq x \leq x_2$. If the constant matrix K is defined by $^1(x_1)K = U(x_1)$, and we set $\Xi^{1*} = \Xi^1 K$, then for arbitrary $a = (a_i)$ we have $\gamma^1{}_{\rho'}(\Xi^{1*}a, U(x_1)a; c) = 0$, $(\rho' = 1, \cdots, k')$, and hence for arbitrary $(\xi^1, a) = (\xi_{\rho'}, a_i)$,

$$\begin{aligned}\gamma^1(\xi^1, U(x_1)a; c) &= \gamma^1(\Xi^{1*}a, U(x_1)a; c) + \gamma^1(\xi^1 - \Xi^{1*}a, 0; c)\\ &\geq \gamma^1(\Xi^{1*}a, U(x_1)a; c),\end{aligned}$$

and the equality sign holds if and only if $\xi^1 = \Xi^{1*}a$. Moreover, by the argument of Bliss [2; pp. 248, 249],

$$2\gamma^1(\Xi^{1*}a, U(x_1)a; c) - \tilde{a}\tilde{U}(x_1)V(x_1)a \geq \tilde{a}a. \tag{10.8}$$

Indeed, in view of the non-singularity of $U^1(x_1)$, the argument of Bliss may be replaced by a simple direct calculation to show that the left-hand member of (10.8) has the value $\tilde{a}a + \tilde{a}a^*$, where a^* is defined by $U^1(x_1)a^* = U^2(x_1)a$, and hence $\tilde{a}a^* \geq 0$. Consequently, (8.9) holds if (ξ^1, a) is a non-null set.

a

), V

ar on

be the corre

(x) of (8.6). $\eta(x)$ is

$x_1 \leq x \leq x_2,$

)). The so-called Clebsch

ty

$UVh)'$,

$u(x) = U(x)h'(x)$.

s an admissible set satisfying (8.2), (8.3),

(10. $J_2(\xi, \eta) =$ $\xi^1, h(x_1) ; c ; x_1) + 2\Gamma^2(\xi^2, h(x_2) ; c ; x_2)$

$$+ \int_{x_1}^{x_2} \tilde{u} R u \, dx.$$

Since (8.2) holds we have $\phi(x)u(x) \equiv 0$, and from III′ it [illegible] that the integral in (10.12) is non-negative, and equal to zero if and only [illegible] $(x) \equiv 0$; that is, if and only if $h'(x) \equiv 0$. Moreover, by (8.9) and (8[illegible]), the first two terms of the right member of (10.12) are non-negative, and are equal to zero if and only if $(\xi^1, h(x_1)) = (0, 0)$ and $(\xi^2, h(x_2)) = (0, 0)$, respectively. That is, $J_2(\xi, \eta) \geq 0$ for arbitrary admissible sets $(\xi, \eta(x))$ satisfying (8.2), (8.3), and the equality sign holds only if $(\xi, \eta(x))$ is null.

11. Results related to Theorems 8.1 and 8.2. If B_2 involves separated end-conditions, satisfies the non-singularity condition, and $J_2(\xi, \eta) \geq 0$ for arbitrary admissible $(\xi, \eta(x))$ satisfying (8.2), (8.3); then if $(\xi, \eta(x))$ is such a set for which $J_2(\xi, \eta) = 0$ it is well-known that there exists a $\zeta(x)$ such that η, ζ is a solution of (8.6); moreover, there are constants $d_{\nu'}$, $d_{\nu''}$, $(\nu' = 1, \cdots, p'; \nu'' = p' + 1, \cdots, p)$, with which $\xi^1, \eta(x_1)$, $\zeta(x_1)$ and $\xi^2, \eta(x_2)$, $\zeta(x_2)$ satisfy the "transversality conditions"[6]

---

[6] See, for example, Bliss [2; p. 232]; indeed, the transversality conditions (11.1) will involve only $p - r_1$ parameters $d_{\nu'}$, $d_{\nu''}$ if the problem B_2 is abnormal or order r_1.

$\eta(x_2)$

$= \gamma^2_{i2}(\xi$

$(\rho' =$

On the other

$J_2(\xi, \eta) > 0$ for arbitrary

(8. 3), the definiteness of J_2

$U(x)$, $V(x)$ and c are deter

quadratic forms (8. 9), (8. 1

that

$$(11.2)\qquad \begin{aligned} &2\Gamma^1(\xi^1, a; \\ &2\Gamma^2(\xi^2, b; \end{aligned}$$

Now, as already pointed out in Section **3**, in view $l \geq 0$ such that for $2\omega(x, \eta, \eta'; l) = 2\omega(x, \eta, \eta'$ re-sponding matrix $R(x; l) = R(x) + l\tilde{\phi}(x)\phi(x)$ positive on $x_1 \leq x \leq x_2$. Moreover, the replacement of $\omega(x, \eta,$ by $\omega(x, \eta, \eta'; l)$ leaves unaltered the coefficients of (8. 6); in particular, solutions $\eta_i = U_{ij}(x)$, $\zeta_i = V_{ij}(x)$ of these equations determined in Theorem 8. 1, and the corresponding liers $\mu_a = \mu_{aj}(x)$, are unaltered by this substitution. If u is defined in (10. 11), then there is a $\kappa'' > 0$ such that $\tilde{u}R(x; l)u \geq \kappa'' \| h'(x) \|^2$ on $x_1 \leq x \leq x_2$. Now by elementary inequalities,

$$\begin{aligned} \| h(x) \|^2 &\leq 2 \| h(x_1) \|^2 + 2 \| \int_{x_1}^{x} h'(t)\, dt \|^2, \\ &\leq 2 \| h(x_1) \|^2 + 2(x - x_1) \int_{x_1}^{x_2} \| h'(t) \|^2\, dt, \end{aligned}$$

and hence

$$\int_{x_1}^{x_2} \| h(x) \|^2 dx \leq 2(x_2 - x_1) \| h(x_1) \|^2 + (x_2 - x_1)^2 \int_{x_1}^{x_2} \| h'(t) \|^2\, dt.$$

Finally, there exist positive constants κ_2, κ_3, dependent only upon $U(x)$, $U'(x)$, such that $\| \eta(x) \|^2 \leq \kappa_2 \| h(x) \|^2$, $\| \eta'(x) \|^2 \leq \kappa_3 (\| h(x) \|^2 + \| h'(x) \|^2)$, $(x_1 \leq x \leq x_2)$. Considering (10. 10), (10. 12) for the corresponding $\Omega(x, \eta, \eta', \mu; l)$, a simple combination of the above inequalities results in the existence of a positive constant κ such that for arbitrary admissible $(\xi, \eta(x))$,

$$(11.3)\qquad \begin{aligned} &2\gamma^1(\xi^1, \eta(x_1); c) + 2\gamma^2(\xi^2, \eta(x_2); c) \\ &\qquad + \int_{x_1}^{x_2} 2\Omega(x, \eta, \eta', \mu(x; \eta); l)\, dx \geq \kappa I(\xi, \eta). \end{aligned}$$

$\eta(x_2)$

end-conditions, satisfies III′
x)) satisfying (8. 2), (8. 3),
satisfying (8. 2), (8. 3), *for*
$\zeta(x)$, $d_{\nu'}$, $d_{\nu''}$ is a solution of
rsality conditions (11. 1), *or*
at (11. 3) *holds for arbitrary*
r case,

$$J(\xi, \eta) \geq \quad (\xi, \eta)$$

for *$\eta(x)$) satisfying* (8. 2), (8. 3).

If B_2 involves s ted end-conditions, satisfies III′, and $J_2(\xi, \eta) > 0$ for arbitrary non-null admissible $(\xi, \eta(x))$ satisfying (8. 2), (8. 3), then for $U(x)$, $V(x)$, c determined as in Theorem 8. 1 there is a > 0 such that $U(x)$ remains non-singular on $x_1 - \delta_0 \leq x \leq x_2 + \delta_0$, while i $x_1 | < \delta_0$, $| X_2 - x_2 | < \delta_0$ the inequalities $2\Gamma^1(\xi^1, a; c; X_1) \geq \kappa_1(\| \xi$ $+ \| a \|^2)$, $2\Gamma^2(\xi^2, b; c; X_2) \geq \kappa_1(\| \xi^2 \|^2 + \| b \|^2)$, corresponding to (11. 2), hold for a suitable positive κ_1. Now if $\eta_i(x)$, $(i = 1, \cdots, n)$, are arbitrar absolutely continuous functions on $X_1 \leq x \leq X_2$, the corresponding functions $h_i(x)$ defined by (10. 9) are also absolutely continuous on this interval. If $\mu_a = \mu_a(x; \eta)$ are defined by (10. 11), and $\pi_i = \pi_i(x; \eta) = U'_{ij}(x) h_j(x)$, clearly the integral

$$(11.6) \qquad J^*_2(\eta; X_1, X_2) \equiv 2 \int_{X_1}^{X_2} \{\Omega(x, \eta, \pi, \mu) + (\eta'_i - \pi_i) \Omega_{\pi_i}(x, \eta, \pi, \mu)\} dx$$

exists. Indeed (11. 6) is the Hilbert invariant integral for the problem B_2 in the field determined by the mutually conjugate solutions $\eta_i = U_{ij}(x)$, $\zeta_i = V_{ij}(x)$ of (8. 6). Moreover, since $\eta'_i = \pi_i + u_i$, where u is defined by (10. 11), it is a consequence of the identity (10. 10) that

$$J^*_2(\eta; X_1, X_2) = \tilde{h}(x) \tilde{U}(x) V(x) h(x) \Big|_{X_1}^{X_2} .$$

Consequently, we have the following result.

for a

e exis

remains no

ompon s of $\eta(x)$ are

$|X_1 - x_1| < \delta_0$, $|X_2 - x$

$$(11.7)\quad \gamma^1(\xi^1,\eta(X_1);c)$$

$$\geq \kappa_1(\|\xi^1\|^2 +$$

with $h(x)$ defined by (10.9)

It is to be remarked

$J^*_2(\eta;X_1,X_2)$ is equal to Φ

hence this coefficient is zero

This fact will be utilized in t ng s

12. A problem involving isoperimetric con B_2 not involving separated end-conditions, results analo to those of Theorems 11.1 and 11.2 may be obtained by reducing such a lem, by a well-known type of transformation (see, for example, Bliss [2], Secs. 69, 88), to one involving se ed end-conditions. For brevity, we consider immediately a more gen roblem involving isoperimetric conditions. The notation B_{2I} will be ass ed to the variational problem involving the quadratic functional (8.1) subject to (8.2), (8.3), and the isoperimetric conditions

$$(12.1)\quad X(\xi,\eta;x_1,x_2) = \chi_s(\xi,\eta(x_1),\eta(x_2)) + \int_{x_1}^{x_2}(\phi^*_{sj}(x)\eta'_j + \theta^*_{sj}\eta_j)\,dx = 0, \qquad (s=1,\cdots,q),$$

where the χ_s are linear forms in ξ_ρ, $\eta_i(x_1)$, $\eta_i(x_2)$ with real coefficients, and the functions $\phi^*_{sj}(x)$, $\theta^*_{sj}(x)$ are real-valued continuous functions on an interval containing $x_1 \leq x \leq x_2$ in its interior.

The Euler-Lagrange equations for B_{2I} are

$$(12.2)\quad (d/dx)\{\Omega_{\pi_i}(x,\eta,\eta',\mu) + e_s\phi^*_{si}\} - \{\Omega_{\eta_i}(x,\eta,\eta',\mu) + e_s\theta^*_{si}\} = 0,$$
$$\Phi_\alpha(x,\eta,\eta') = 0,$$

where the constants e_s, $(s=1,\cdots,q)$, are "isoperimetric parameters." The corresponding transversality conditions are

$si] = 0,$

$X_2 = 0$

$= 1, \cdots, q)$, are defined as

$+ \theta^*_{sj}\eta_j)\, dx, \qquad x_1 \leq x \leq x_2,$

m $\mathbf{B}_2$ in $(\xi, \boldsymbol{\eta}) = (\xi_\rho, \boldsymbol{\eta}_\sigma)$
onal

$$\int_{x_1}^{x_2} 2\omega(x, \eta, \eta')\, dx$$

su homogeneous differ tial equations

$$(12. \quad \Phi_\alpha(x, \eta, \eta' \quad 0,$$

$$(12.7) \qquad \eta^1{}_i{}' = 0, \qquad \eta^2{}_s{}' + \phi^*_{sj}\eta'_j + \theta^*_{sj}\eta_j = 0,$$

and the linear homogeneous end-conditions

$$(12.8) \qquad \Psi_\nu(\xi, \eta(x_1), \eta^1(x_1)) = 0,$$

$$(12.9) \qquad \eta_i(x_2) - \eta^1{}_i(x_2) = 0, \qquad \eta^2{}_s(x_2) = 0,$$
$$\chi_s(\xi, \eta(x_1), \eta^1(x_1)) + \eta^2{}_s(x_1)) = 0.$$

The problem $\mathbf{B}_2$ clearly involves separated end-conditions; moreover, for $\mathbf{B}_2$ each of the conditions of non-singularity, III, or III′, is equivalent to the corresponding condition for the initial problem B_2, or for the isoperimetric problem B_{2I}. For a non-singular problem the canonical form of the Euler-Lagrange equations for $\mathbf{B}_2$ will be denoted by

$$(12.10) \qquad \boldsymbol{\eta}' = \boldsymbol{A}(x)\boldsymbol{\eta} + \boldsymbol{B}(x)\boldsymbol{\zeta}, \qquad \boldsymbol{\zeta}' = \boldsymbol{C}(x)\boldsymbol{\eta} - \tilde{\boldsymbol{A}}(x)\boldsymbol{\zeta}.$$

If (8.1) is positive for arbitrary non-null admissible $(\xi, \eta(x))$ satisfying (8.2), (8.3), (12.1), then (12.5) is positive for arbitrary non-null admissible $(\xi, \boldsymbol{\eta}(x)) = (\xi, \eta(x), \eta^1(x), \eta^2(x))$ satisfying (12.6), (12.7), (12.8), (12.9). In this case we shall denote by $\boldsymbol{\eta}_\sigma = \boldsymbol{U}_{\sigma\tau}(x)$, $\boldsymbol{\zeta}_\sigma = \boldsymbol{V}_{\sigma\tau}(x)$. $(\tau = 1, \cdots, \boldsymbol{n} = 2n + p)$, a conjugate system of solutions of (12.10) with

ion
), (8
that
he differenti
rst 2n nditions of (12.

(12.11) $\boldsymbol{\eta}($

the equation

(12.12) $\boldsymbol{\eta}(x) =$

corresponding to (10.9), d
admissible components. F
(12.10) there are correspon
with $\boldsymbol{U}_{\sigma\tau}(x)$ satisfy the ord
$\mathbf{B}_2$; in particular, $\boldsymbol{\mu}^1_{i\tau}(x) \equiv$ an
constants. In terms of he Euler equations for B
$\mu_a = \boldsymbol{\mu}_{a\tau}(x)$, $e_s = \boldsymbol{V}_{2n+s,\tau}$ are solutions of (12. $x)$
we shall denote by $\boldsymbol{\mu}_a = \boldsymbol{\mu}_a(x;\eta)$ the values $\boldsymbol{\mu}_{a\tau}$ $_\tau(x)$, where is
defined by (12.11), (12.12). In view of the auxi differential equations
and end-conditions for $\mathbf{B}_2$ satisfied by an $\boldsymbol{\eta}$ of the form (12.11), and since
$\| \eta(x) \| \leq \|$ $\|$, $\| \eta'(x) \| \leq \| \boldsymbol{\eta}'(x) \|$, the following result for B_{2I} is an
immediat equence of Theorem 11.1 for $\boldsymbol{B}_2$.

THEOREM 12.1. *If* B_{2I} *satisfies* III′, *and* $J_2(\xi, \eta) \geq 0$ *for arbitrary admissible* $(\xi, \eta(x))$ *satisfying* (8.2), (8.3), (12.1), *then either there is a non-null admissible* $(\xi, \eta(x))$ *satisfying* (8.2), (8.3), (12.1) *for which* $J_2(\xi, \eta) = 0$, *and which satisfies the differential equations* (12.2) *and transversality conditions* (12.3) *with associated* $\mu_a(x)$, e_s, d_ν, *or there exist constants* $c \geq 0$, $l \geq 0$, $\kappa > 0$ *such that for arbitrary admissible* $(\xi, \eta(x))$,

$$\begin{aligned} &2\gamma(\xi, \eta(x_1), \eta(x_2)) + c(\| \Psi(\xi, \eta(x_1), \eta(x_2)) \|^2 + \| \mathrm{X}(\xi, \eta; x_1, x_2) \|^2) \\ &\qquad + \int_{x_1}^{x_2} 2\Omega(x, \eta, \eta', \boldsymbol{\mu}(x;\eta); l)\, dx \geq \kappa I(\xi, \eta), \end{aligned} \tag{12.13}$$

where

$$\begin{aligned} I(\xi, \eta) = \| \xi \|^2 + \| \eta(x_1) \|^2 + \| \eta(x_2) \|^2 + \\ + \int_{x_1}^{x_2} (\| \eta \|^2 + \| \eta' \|^2)\, dx. \end{aligned} \tag{12.14}$$

In particular, in this latter case, $J_2(\xi, \eta) \geq \kappa I(\xi, \eta)$ *for arbitrary admissible* $(\xi, \eta(x))$ *satisfying* (8.2), (8.3), (12.1).

for $\mathbf{B}_2$.

$(\xi, \eta) > 0$ *f*

), (12.1),

) $= \| \boldsymbol{U}_{\sigma\tau}(x) \|$ *of Theorem*

, and if the com-

, where $| X_1 - x_1 |$

$\| \mathbf{X}(\xi, \eta; X_1, X_2) \|^2)$,

), *is defined by*

$$= \eta_i(x), \quad {}_{n+i,\tau}(x) \boldsymbol{h}_\tau(x) = \eta_i(X_2),$$

$$(x) = \int_x^{X_2} (\phi^*_{sj}\eta'_j + \theta^*_{sj}\eta_j)\, dx,$$

with $\boldsymbol{\pi}_i = \boldsymbol{\pi}_i(x; \eta) = {}_{\tau}(x) \boldsymbol{h}_\tau(x)$, $\boldsymbol{\mu}_a = \boldsymbol{\mu}_a(x; \eta) = \boldsymbol{\mu}_{a\tau}(x) \boldsymbol{h}_\tau(x)$, *where* $\boldsymbol{\mu}_{a\tau}(x)$ *are corresponding multipliers for the* $\boldsymbol{U}_{i\tau}(x)$, *and*

$$(12.17) \quad \boldsymbol{J}^*_2(\eta; X_1, X_2) = 2 \int_{X_1}^{X_2} \{\Omega(x, \eta, \boldsymbol{\pi}, \boldsymbol{\mu}) + (\eta'_i - \boldsymbol{\pi}_i)\Omega \quad \boldsymbol{\pi}, \boldsymbol{\mu})\} dx.$$

13. Concluding remarks. Theorem 12.1 is of use in the consideration of self-adjoint boundary problems. For example, in the boundary problem considered by the author [11] this theorem enables one to establish the existence of characteristic values and corresponding characteristic solutions possessing definitive extremizing properties. Such a proof is simpler in detail than that employed in [11], which utilized the Green's matrix and generalized Green's matrix for differential systems.

For ordinary problems of Bolza involving separated end-conditions Theorem 11.1 has the effect of allowing one to disregard in a certain sense the auxiliary differential equations and end-conditions. More precisely, suppose $E: y_i(x)$, $\lambda_0 = 1$, $\lambda_a(x)$, $x_1 \leq x \leq x_2$, is an extremal for such a problem of Bolza involving (2.1) subject to (2.3) and (2.4), and satisfying with constants e_ν the multiplier rule and conditions III′, IV′. If g is replaced by $g + c\psi_\nu\psi_\nu$, and f by $f + (\mu_a(x; y - y(x)) + l\phi_a)\phi_a$, where c, l and $\mu_a(x; \eta)$ are as in Theorem 11.1, then the resulting minimum problem has

ons
is an ex
e seco
f of this res
sufficien proof for such a
no auxiliary differential
fundamental prope s
have been used in its deriva
ever, in a manner similar to
of [8] for his formulation o
Indeed, for such a problem
of the integrand function o
given problem from the suffic
no side differential equations
separated end-conditions, or fo erm
of Theorem 12.1 does not lead in a correspondi
the original type for which the effect of the s
and end-conditions is eliminated for the second ion. How is
property is enjoyed by an equivalent formulation e general problem of
Bolza, which may be described briefly as a hybrid of that employed by
Hestenes and ne here used.

NORTH N UNIVERSITY.

REFERENCES.

1. G. A. Bliss, "The transformation of Clebsch in the calculus of variations," *Proceedings of the International Mathematical Congress held in Toronto*, vol. 1 (1924), pp. 589-603.

2. ———, *Lectures on the calculus of variations*, University of Chicago Press, 1946.

3. M. R. Hestenes, "Sufficient conditions for the problem of Bolza in the calculus of variations," *Transactions of the American Mathematical Society*, vol. 36 (1934), pp. 793-818.

4. M. R. Hestenes and W. T. Reid, "A note on the Weierstrass condition in the calculus of variations," *Bulletin of the American Mathematical Society*, vol. 45 (1939), pp. 471-473.

ety, vol. 6

of Bolza in nonpara-
tical Society, vol. 62 (1947), pp.

r konjugierten Punkte beim Prob-
. 57 (1927), pp. 243-257.
ndlungen aus dem Mathematischen

associated with the calculus of
4 (1932), pp. 769-790.

methods for the problem of Bolza
nts of *ic*, vol. 38 (1937), pp. 662-678.

sion proof of sufficient nditions for the non-para-
tions of the American athematical Society, vol. 42

p. 183-190.

14. ———, "A em on quadratic forms," *Bulletin of the American Mathematical Society*, vol. 44 (1938), pp. 437-440.

15. ———, "Isoperimetric problems of Bolza in non-parametric form," *Duke Mathematical Journal*, vol. 5 (1939), pp. 675-691.

16. ———, "A matrix differential equation of Riccati type, can Journal of Mathematics*, vol. 68 (1946), pp. 237-246.

17. ———, "Addendum" to preceding paper, *American Journal of Mathematics*, vol. 70 (1948), p. 460.

CALCUTTA UNIVERSITY
CENTRAL LIBRARY